From Markov Chains to Gibbs Fields.

Yevgeniy Kovchegov
Department of Mathematics,
Oregon State University
Corvallis, OR
97331-4605, USA
kovchegy@math.oregonstate.edu

Contents

1 Markov Chains: Time reversibility, recurrence and ergodicity. 5
- 1.1 Markov chains, detailed balance condition and reversibility. 5
  - 1.1.1 Markov chains, stationary distributions and reversibility via traffic flows. 5
  - 1.1.2 Recurrence. 7
- 1.2 Recurrence and electrical networks. 8
  - 1.2.1 Finite electrical networks. 9
  - 1.2.2 Infinite electrical networks. 14
- 1.3 Martingales and Lyapunov functions 16
  - 1.3.1 Martingales and harmonic functions 16
  - 1.3.2 Stopping theorem 17
  - 1.3.3 Lyapunov functions 17
  - 1.3.4 Example: 1-D Random Walk in Random Environment. 18
  - 1.3.5 Martingale convergence theorem 20
- 1.4 Homework #1 20

2 Long Run Behavior of Stochastic Processes. 21
- 2.1 Coupling method, convergence rates via coupling. 21
- 2.2 Second largest eigenvalue. 22
  - 2.2.1 Spectral gap and spectral theorem. 24
  - 2.2.2 Relaxation times. 28
- 2.3 Mixing times. 30
  - 2.3.1 Strong stationary times. 30
  - 2.3.2 Card shuffling examples and cutoff asymptotics. 32
  - 2.3.3 Mixing times via coupling. 33
  - 2.3.4 Mixing and relaxation times. 35
- 2.4 Long run behavior of continuous time processes. 36
  - 2.4.1 Example: Ising model and Glauber dynamics. 36
- 2.5 Homework #2. 36

3 Gibbs Fields 39
- 3.1 Markov random fields. 39
  - 3.1.1 Gibbs-Markov equivalence. 40
- 3.2 Ising model and bond percolation. 42
  - 3.2.1 Spin systems 42

3.2.2 Phase transition. 42
3.2.3 Criticality. 42
3.2.4 Two dimensional bond percolation: $p_c = \frac{1}{2}$. 42
3.3 Monte Carlo Markov Chain method, Metropolis Algorithm and Gibbs sampler. 42

Chapter 1

Markov Chains: Time reversibility, recurrence and ergodicity.

1.1 Markov chains, detailed balance condition and reversibility.

We begin with n cities (states) and matrix of one-step transition probabilities $P = \{p(i,j)\}$. If μ population distribution (in fractions) among the n cities, then after one unit of time, the population is distributed according to μP (μ is an n-dimensional vector).

We recall Chapman-Kolmogorov theorem,

$$p_n(i,j) = \sum_k p_m(i,k) p_{n-m}(k,j).$$

Thus after n units of time, μP^n is our new distribution.

1.1.1 Markov chains, stationary distributions and reversibility via traffic flows.

The stationary distribution π is defined so that

$$\pi P = \pi \qquad \Leftrightarrow \qquad \sum_i \pi(i) p(i,j) = \pi(j).$$

Thus $\sum_i \pi(i) p(i,j) = \pi(j) \sum_i p(j,i)$, and for any city j,

$$\sum_{i \neq j} \pi(i) p(i,j) = \pi(j) \sum_{i \neq j} p(j,i).$$

Thus in terms of population traffic, the inflow to the city j is equal to outflow from j, for each j. Thus the distribution of population stays unchanged.

The following are the *detailed balance* conditions (d.b.c.) also called *time reversibility*:

$$\pi(i) p(i,j) = \pi(j) p(j,i).$$

In other words, for every two cities i and j the traffic in between them is balanced, i.e. the traffic flow from i to j is equal to the traffic flow from j to i. It is easy to see that if d.b.c. are satisfied, the population distribution will not change with time.

Exercise. Show mathematically that if π satisfies the detailed balance conditions, then it is stationary, i.e. π satisfies d.b.c. $\Rightarrow \pi$ is stationary.

Exercise. Give a counterexample in order to prove: π is stationary $\not\Rightarrow \pi$ satisfies d.b.c.

A Markov chain is said to be time reversible if it has a stationary distribution π such that the probability of starting at i and going to j in n steps via any given path

$$i \to i_1 \to i_2 \to \cdots \to i_{n-1} \to j$$

is the same as if we were to start at j and go in the reversed order to i:

$$j \to i_{n-1} \to \cdots \to i_2 \to i_1 \to i.$$

In other words, if we reverse the direction of time axis, the same trajectory will have the same probability. Time reversibility can be used to extend a stationary chain to negative times.

Theorem 1. *The detailed balance condition implies reversibility.*

Proof:

$$\begin{aligned}
\mathbf{Prob}[i \to i_1 \to i_2 \to \cdots \to i_{n-1} \to j] &= \pi(i)p(i,i_1)p(i_1,i_2)\cdot\cdots\cdot p(i_{n-1},j) \\
&= p(i_1,i)\pi(i_1)p(i_1,i_2)\cdot\cdots\cdot p(i_{n-1},j) \\
&= p(i_2,i_1)p(i_1,i)\pi(i_2)p(i_2,i_3)\cdot\cdots\cdot p(i_{n-1},j) \\
&\cdots \\
&= \pi(j)p(j,i_{n-1})\cdot\cdots\cdot p(i_2,i_1)p(i_1,i) \\
&= \mathbf{Prob}[j \to i_{n-1} \to \cdots \to i_2 \to i_1 \to i].
\end{aligned}$$

□

Example. *Birth-and-death chain* States: $0, 1, \ldots,$.
Probabilities: $p(0,1) = p_0 = 1 - p(0,0)$ and

$$p(j, j+1) = p_j \quad \text{and} \quad p(j, j-1) = q_j = 1 - p_j \quad \text{for any} \quad j = 1, 2, \ldots \quad .$$

Now, from reversibility we have $\pi(1) = \frac{p_0}{q_1}\pi(0)$,

$$\pi(2) = \frac{p_1}{q_2}\pi(1) = \frac{p_0 p_1}{q_1 q_2}\pi(0),$$

$$\pi(3) = \frac{p_2}{q_3}\pi(2) = \frac{p_0 p_1 p_2}{q_1 q_2 q_3}\pi(0),$$
$$\dots$$
$$\pi(j) = \frac{p_{j-1}}{q_j}\pi(j-1) = \frac{p_0 p_1 \dots p_{j-1}}{q_1 q_2 \dots q_j}\pi(0),$$

and so on. We find $\pi(0)$ from $\pi(0)+\pi(1)+\pi(2)+\cdots = 1$, but we can do it only if the series

$$\sum_j \frac{p_0 p_1 \dots p_{j-1}}{q_1 q_2 \dots q_j} < \infty \ .$$

Example. *Diamond plus birth-and-death.* States: $0, 1, 1^*, \dots,$.
Probabilities: $p(0,1) = p_0 = 1 - p(0,1^*) = 1 - p_0^*$,

$$p(1,2) = p_1 = 1 - p(1,0), \qquad p(1^*,2) = p_1^* = 1 - p(1^*,0),$$

$$p(2,3) = p_2, \qquad p(2,1) = q_2, \qquad p(2,1^*) = q_2^* \qquad \text{where } p_2 + q_2 + q_2^* = 1$$

and $p(j, j+1) = p_j = 1 - p(j, j-1)$ for $j = 3, 4, \dots$.

1.1.2 Recurrence.

Let T_x be the first time city (site) x is visited.

Definition. State x is said to be *recurrent* if

$$\mathbf{Prob}[T_x < \infty \mid X_0 = x] = 1,$$

and otherwise it is called *transient*.

Definition. A recurrent state x is said to be *positive* recurrent if

$$\mathbf{E}[T_x \mid X_0 = x] < \infty \ ;$$

x is said to be *null* recurrent if

$$\mathbf{E}[T_x \mid X_0 = x] = \infty \ .$$

If the Markov chain is irreducible and aperiodic, and has a stationary distribution π, then every site x is positive recurrent and

$$\mathbf{E}[T_x \mid X_0 = x] = \frac{1}{\pi(x)} \ .$$

Let $\mathcal{F}_m = \mathcal{F}(X_0, X_1, ..., X_m)$ denote the history of the process up to time m.

Definition. A random variable τ is a *stopping time* if for any $m \geq 0$,

$$\{\tau \leq m\} \in \mathcal{F}_m.$$

In other words knowing the trajectory of the process up to time m is sufficient to determine whether $\{\tau \leq m\}$ occurred.

Thus T_x is a stopping time.

Example. *Why $\mathcal{F}_m$ is a σ-algebra?*

Consider a collection of independent Bernoulli random variables $\xi_1, \xi_2, ...$ with states 0 and 1, and a Markov chain $\{X_n\}_n$,

$$\text{where} \quad X_0 = 0 \quad \text{and} \quad X_n = \xi_0 + \xi_1 + \cdots + \xi_n.$$

The history of the process up to time m can be represented by a random binary number $0.\xi_1\xi_2\xi_3 \ldots \xi_m = \sum_{j=1}^{m} \frac{\xi_j}{2^j}$, or rather by the half-open interval

$$\left[0.\xi_1\xi_2\xi_3 \ldots \xi_m, \quad 0.\xi_1\xi_2\xi_3 \ldots \xi_m + \frac{1}{2^m}\right).$$

We begin with the $[0,1)$ interval. We split it in two halves. If $\xi_1 = 0$ we select the left subinterval $[0, \frac{1}{2})$, and if $\xi_1 = 1$ we select the right subinterval $[\frac{1}{2}, 1)$. The selected interval is again being split in two: We take the left subinterval if $\xi_2 = 0$, and the right subinterval if $\xi_2 = 1$. And so on.

Now after m iterations we arrive with one of 2^m small intervals

$$\left[0, \frac{1}{2^m}\right), \quad \left[\frac{1}{2^m}, \frac{2}{2^m}\right), \quad \left[\frac{2}{2^m}, \frac{3}{2^m}\right), \ldots, \quad \left[\frac{2^m - 1}{2^m}, 1\right).$$

Each interval represents one of 2^m possible histories. Thus $\mathcal{F}_m$ is the σ-algebra generated by the above intervals, i.e. any event A that is entirely determined by the history $\mathcal{F}_m$ can be represented as a subset of $[0,1)$ that is a union of some of these intervals. We write $A \in \mathcal{F}_m$.

Here the stopping time could be the first time T_2 the Markov chain $\{X_n\}_n$ hits 2. One can check that the event $\{T_2 \leq 3\}$ corresponds to $0.\xi_1\xi_2\xi_3$ being equal to either 0.011, 0.101, 0.110 or 0.111 in base two. Thus

$$\{T_2 \leq 3\} = \left[\frac{3}{8}, \frac{1}{2}\right) \bigcup \left[\frac{5}{8}, \frac{3}{4}\right) \bigcup \left[\frac{3}{4}, \frac{7}{8}\right) \bigcup \left[\frac{7}{8}, 1\right) \in \mathcal{F}_3 \; .$$

Definition. *Filtration of σ-algebras:* $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots$.

1.2 Recurrence and electrical networks.

G.Lawler; Y.Peres; Aldous and Fill;

[Doyle and Snell] is the primary reference (posted on the web). I will rephrase many lines from Yuval Peres.

1.2.1 Finite electrical networks.

Yuval Peres: *While electrical networks are only a different language for reversible Markov chains, the electrical point of view is useful because of the insight gained from the familiar physical laws of electrical networks.*

Here each edge e of a finite connected graph (*network*) G has a **conductance** value c_e attached to it. We also define the resistance of an edge as the reciprocal value $r_e = \frac{1}{c_e}$.

If we define $\pi(x) = \sum_{y:x\sim y} c_{xy}$ and $p(x,y) = \textbf{Prob}[X_{n+1} = y \mid X_n = x] = \frac{c_{xy}}{\pi(x)}$ (where $x \sim y$ means sites x and y are connected by an edge of G) to be respectively a probability distribution and transition probabilities. Then the Markov chain will be reversible (with respect to $\pi(\cdot)$, that can be rescaled by $2\sum_e c_e$ to be a probability distribution):

$$\pi(x)p(x,y) = c_{xy} = \pi(y)p(y,x).$$

And vice-versa, if the Markov chain is reversible with stationary distribution π, the conductance c_{xy} can be defined as above. Thus the reversibility allowes us to represent the Markov chain $\{X_n\}$ as a random walk on a *weighted* graph with weights $\{c_e\}_e$.

Example. *Simple random walk on G.* There $c_{x,y} = 1$ for $x \sim y$ and therefore $p_{x,y} = \frac{1}{\deg(x)}$ for $x \sim y$.

Example. *Chess knight random walk.*
Aldous and Fill: *Start a knight at a corner square of an otherwise-empty chessboard. Move the knight at random, by choosing uniformly from the legal knight-moves at each step. What is the mean number of moves until the knight returns to the starting square?* Solution: we let the conductance between sites that are knight-move away from each other to be $= 1$. Thus $\sum_e c_e$ will be equal to the number of edges in the graph $|\mathcal{E}| = 168$ and $\pi(x) = \frac{\deg(x)}{2|\mathcal{E}|}$. Since we began the knight walk at a corner site v with $\deg(v) = 2$,

$$\pi(v) = \frac{1}{|\mathcal{E}|} \quad \text{and} \quad \mathbf{E}[T_v \mid X_0 = v] = \frac{1}{\pi(v)} = |\mathcal{E}| = 168\ .$$

We fix two nodes (vertices) **a** and **b** to be the two poles of the electrical network. Let T_a be the first time the Markov chain $\{X_n\}$ arrives to **a**, and similarly, let T_b be the first time the Markov chain arrives to **b**. Then $\{T_a < T_b\}$ is the event that the Markov chain hits **a** before it hits **b**. Suppose we are given the **voltage** values, $V_a > V_b$, at **a** and **b**. Then for the rest of the vertices, we define the **voltage** as

$$V_x = V_b + \textbf{Prob}[\{T_a < T_b\} \mid X_0 = x] \cdot (V_a - V_b).$$

Here $\textbf{Prob}[\{T_a < T_b\} \mid X_0 = x] = \frac{V_x - V_b}{V_a - V_b}$ the expression that one obtains when $h(x) = V_x$ is a harmonic function of x and $h(X_n)$ is a martingale. Please keep this connection in mind when we soon cover some of the martingale theory.

Indeed $h(x) = V_x$ is *harmonic*:

$$\sum_y p(x,y)h(y) = h(x)$$

since $\sum_y p(x,y)\mathbf{Prob}[\{T_a < T_b\} \mid X_0 = y] = \mathbf{Prob}[\{T_a < T_b\} \mid X_0 = x]$.

Now, given a voltage V on the network, the **current flow** associated with V is defined on oriented edges by

$$I(\overrightarrow{xy}) = \frac{V_x - V_y}{r_{xy}}.$$

The above definition is really just the **Ohm's law**: $r_{xy}I(\overrightarrow{xy}) = V_x - V_y$.

The current flow satisfies the **cycle law**: If $c_{x_0x_1} \neq 0$, $c_{x_1x_2} \neq 0$, $c_{x_2x_3} \neq 0, \ldots, c_{x_{n-2}x_{n-1}} \neq 0$ and $c_{x_{n-1}x_n} \neq 0$, where $x_n = x_0$, then

$$\sum_{k=1}^{n} r_{x_{i-1}x_i} I(\overrightarrow{x_{i-1}x_i}) = 0.$$

In general, a **flow** θ from **a** to **b** is an antisymmetric function on oriented edges (i.e. $\theta(\overrightarrow{xy}) = -\theta(\overrightarrow{yx})$) that obeys

Kirchhoff's node law: $$\sum_{w:w\sim v} \theta(\overrightarrow{vw}) = 0 \quad \text{for each } v \notin \{\mathbf{a},\mathbf{b}\}.$$

The *strength* of an arbitrary flow θ is measured at one of the poles **a**:

$$\|\theta\| = \sum_x \theta(\overrightarrow{\mathbf{a}x}).$$

Note: The current flow I is the only flow from **a** to **b** of strength $\|I\|$ that satisfies the cycle law. How: Consider the difference flow $\theta - I$. Node law plus cycle law imply $\theta - I$ is zero along any path from **a** to **b**.

Finally, we define the **effective resistance**,

$$\mathcal{R}(\mathbf{a} \leftrightarrow \mathbf{b}) = \frac{V_a - V_b}{\|I\|} \ ,$$

and the **effective conductance** $\mathcal{C}(\mathbf{a} \leftrightarrow \mathbf{b}) = \frac{1}{\mathcal{R}(\mathbf{a}\leftrightarrow\mathbf{b})}$.

Now, the probability of leaving **a** an getting to **b** before returning back to **a**,

$$\begin{aligned}
\mathbf{Prob}[\mathbf{a} \to \mathbf{b}] &= \sum_x p(\mathbf{a}, x)\mathbf{Prob}[\{T_b < T_a\} \mid X_0 = x] \\
&= \sum_x p(\mathbf{a}, x)\frac{V_a - V_x}{V_a - V_b} \\
&= \frac{1}{\pi(a)(V_a - V_b)} \sum_x c_{\mathbf{a},x}(V_a - V_x) \\
&= \frac{1}{\pi(a)(V_a - V_b)} \cdot \|I\| \\
&= \frac{1}{\pi(a)\mathcal{R}(\mathbf{a} \leftrightarrow \mathbf{b})}
\end{aligned}$$

Thus

$$\mathbf{Prob}[\mathbf{a} \to \mathbf{b}] = \frac{1}{\pi(a)}\mathcal{C}(\mathbf{a} \leftrightarrow \mathbf{b}) \tag{1.1}$$

This will be important for solving recurrence/transience questions.

Now it is time to recall the long forgotten laws of electricity:

Parallel Law. If two vertices, x_1 and x_2 are connected by two different edges, e_1 and e_2 with respective conductance values c_1 and c_2. Then both edges can be replaced with one, with conductance $c_1 + c_2$. Thus the new resistance equals $\frac{1}{(1/r_1+1/r_2)}$.

Series Law. If the vertex x is connected (wit edges of non-zero conductance) only to two vertices, x_1 and x_2, then

$$r_{x_1x_2} := r_{x_1x} + r_{xx_2} \quad \text{and} \quad I(\overrightarrow{x_1x_2}) := I(\overrightarrow{x_1x}) = I(\overrightarrow{xx_2})$$

One can also **glue** vertices that have the same voltage value together into one vertex since the current never flows between vertices with the same voltage.

Example. *Finite birth-and-death chain.* States: $0, 1, \dots, M$.

Transition probabilities: $p(0, 1) = p_0 = 1 - p(0, 0)$,

$$p(j, j+1) = p_j \quad \text{and } p(j, j-1) = q_j = 1 - p_j \quad \text{for all} \quad j = 1, 2, \dots, M-1$$

and $p(M, M-1) = 1$.

Question: Find the probability $\mathbf{Prob}[\mathbf{0} \to \mathbf{M}]$ of departing from zero and hitting m before returning to zero.

Solution: As we have seen, the above Markov chain is reversible. Thus the above conductance/resistance representation works. Namely, $\mathbf{a} = 0$ and $\mathbf{b} = M$, and the effective resistance

$$\mathcal{R}(0 \leftrightarrow M) = \frac{1}{c_{01}} + \frac{1}{c_{12}} + \dots + \frac{1}{c_{M-1M}} = \frac{1}{\pi(0)p_0} + \frac{1}{\pi(1)p_1} + \frac{1}{\pi(2)p_2} + \dots + \frac{1}{\pi(M-1)p_{M-1}}.$$

Now, from reversibility we know that $\pi(1) = \frac{p_0}{q_1}\pi(0)$,

$$\pi(2) = \frac{p_1}{q_2}\pi(1) = \frac{p_0p_1}{q_1q_2}\pi(0),$$

$$\pi(3) = \frac{p_2}{q_3}\pi(2) = \frac{p_0p_1p_2}{q_1q_2q_3}\pi(0),$$

$$\ldots$$

$$\pi(M-1) = \frac{p_{M-2}}{q_{M-1}}\pi(M-2) = \frac{p_0p_1\ldots p_{M-2}}{q_1q_2\ldots q_{M-1}}\pi(0).$$

Thus

$$\mathcal{R}(0 \leftrightarrow M) = \frac{1}{\pi(0)p_0} + \frac{q_1}{\pi(0)p_0p_1} + \frac{q_1q_2}{\pi(0)p_0p_1p_2} + \frac{q_1q_2q_3}{\pi(0)p_0p_1p_2p_3} + \cdots + \frac{q_1q_2\ldots q_{M-1}}{\pi(0)p_0p_1p_2\ldots p_{M-1}}.$$

Hence, by (1.1),

$$\mathbf{Prob}[\mathbf{0} \to \mathbf{M}] = \frac{1}{\pi(0)\mathcal{R}(0 \leftrightarrow M)} = \frac{p_0}{1 + \frac{q_1}{p_1} + \frac{q_1q_2}{p_1p_2} + \frac{q_1q_2q_3}{p_1p_2p_3} + \cdots + \frac{q_1q_2\cdots q_{M-1}}{p_1p_2\cdots p_{M-1}}}.$$

Conservation of Energy: If θ is a flow from $\mathbf{a}$ to $\mathbf{b}$, then

$$(V_a - V_b)\cdot\|\theta\| = \frac{1}{2}\sum_{x,y:x\sim y}(V_x - V_y)\theta(xy).$$

Proof: Use Kirchhoff's node law.

$$\begin{aligned} \frac{1}{2}\sum_{x,y:x\sim y}(V_x - V_y)\theta(\overrightarrow{xy}) &= \frac{1}{2}\left(\sum_x V_x \sum_{y:x\sim y}\theta(\overrightarrow{xy}) + \sum_{x,y:x\sim y} V_y\theta(\overrightarrow{yx})\right) \\ &= \sum_x V_x \sum_{y:x\sim y}\theta(\overrightarrow{xy}) \\ &= V_a\cdot\|\theta\| - V_b\cdot\|\theta\| \end{aligned}$$

□

Thomson's Principle. [Doyle and Snell, p.50; Y.Peres, p.26]
If I is a unit current flow ($\|I\| = 1$) from $\mathbf{a}$ to $\mathbf{b}$, then the energy $\mathcal{E}(I) = \sum_e [I(e)]^2 r_e$ minimizes the energy $\mathcal{E}(\theta) := \sum_e [\theta(e)]^2 r_e$ among all unit flows θ ($\|\theta\| = 1$) from $\mathbf{a}$ to $\mathbf{b}$. Moreover

$$\mathcal{R}(\mathbf{a} \leftrightarrow \mathbf{b}) = \inf\{\mathcal{E}(\theta)\ :\ \theta \text{ is a unit flow from } \mathbf{a} \text{ to } \mathbf{b}\} = \mathcal{E}(I).$$

Proof: Let θ ($\|\theta\| = 1$) be a unit flow from **a** to **b** and let $D(e) = \theta(e) - I(e)$ be the difference flow with $\|D\| = 0$.

$$\begin{aligned}\mathcal{E}(\theta) &= \frac{1}{2}\sum_{x,y:x\sim y}[I(\overrightarrow{xy}) + D(\overrightarrow{xy})]^2 r_{xy}\\ &= \frac{1}{2}\sum_{x,y:x\sim y}[I(\overrightarrow{xy})]^2 r_{xy} + \sum_{x,y:x\sim y} I(\overrightarrow{xy})D(\overrightarrow{xy})r_{xy} + \frac{1}{2}\sum_{x,y:x\sim y}[D(\overrightarrow{xy})]^2 r_{xy}\\ &= \mathcal{E}(I) + \sum_{x,y:x\sim y}(V_x - V_y)D(\overrightarrow{xy}) + \mathcal{E}(D)\\ &= \mathcal{E}(I) + \mathcal{E}(D) \geq \mathcal{E}(I)\end{aligned}$$

by Conservation of Energy law.

Now,

$$\begin{aligned}\mathcal{E}(I) &= \frac{1}{2}\sum_{x,y:x\sim y}[I(\overrightarrow{xy})]^2 r_{xy}\\ &= \frac{1}{2}\sum_{x,y:x\sim y}(V_x - V_y)I(\overrightarrow{xy})\\ &= (V_a - V_b)\cdot\|I\| = V_a - V_b\\ &= \frac{V_a - V_b}{\|I\|} = \mathcal{R}(\mathbf{a}\leftrightarrow\mathbf{b})\end{aligned}$$

$\square$

Remark. Suppose the strength of the current flow $\|I\| = \alpha \neq 1$. For each edge e we divide the corresponding conductance value by α and get new conductance values $c'_e = \frac{c_e}{\alpha}$. The transition probabilities for the Markov chain (the random walk on the weighted graph) will remain the same, as well as the voltage values

$$V_x = V_b + \mathbf{Prob}[\{T_a < T_b\} \mid X_0 = x]\cdot(V_a - V_b).$$

The new current circuit $I'(\overrightarrow{xy}) := c'_{xy}(V_x - V_y) = \frac{1}{\alpha}I(\overrightarrow{xy})$ will have strength $\|I'\| = 1$. The above argument for $\|I\| = \alpha$ shows

$$\mathcal{E}(I) = (V_a - V_b)\cdot\|I\| = \alpha^2\frac{V_a - V_b}{\|I\|} = \alpha^2\mathcal{R}(\mathbf{a}\leftrightarrow\mathbf{b}).$$

Now, by definition,

$$\mathcal{E}(I) = \frac{1}{2}\sum_{x,y:x\sim y} r_{xy}[I(\overrightarrow{xy})]^2 = \frac{1}{2}\sum_{x,y:x\sim y} c_{xy}(V_x - V_y)^2 = \alpha\cdot\frac{1}{2}\sum_{x,y:x\sim y} c'_{xy}(V_x - V_y)^2$$

$$= \alpha\mathcal{E}(I') = \alpha\mathcal{R}'(\mathbf{a}\leftrightarrow\mathbf{b}),$$

where $\mathcal{R}'(\mathbf{a}\leftrightarrow\mathbf{b})$ is the effective resistance with respect to $\{c'_e\}$. Thus $\mathcal{R}'(\mathbf{a}\leftrightarrow\mathbf{b}) = \alpha\mathcal{R}(\mathbf{a}\leftrightarrow\mathbf{b})$. Hence

$$\mathbf{Prob}[\mathbf{a}\to\mathbf{b}] = \frac{1}{\pi(\mathbf{a})\mathcal{R}(\mathbf{a}\leftrightarrow\mathbf{b})} = \frac{1}{\pi'(\mathbf{a})\mathcal{R}'(\mathbf{a}\leftrightarrow\mathbf{b})},$$

where $\pi(\mathbf{a}) = \sum_{x \sim \mathbf{a}} c_{x\mathbf{a}}$ and $\pi'(\mathbf{a}) = \sum_{x \sim \mathbf{a}} c'_{x\mathbf{a}} = \frac{1}{\alpha}\pi(\mathbf{a})$.

Consider adding an extra edge to the network which is not incident to $\mathbf{a}$. How will this affect $\mathbf{Prob}[\mathbf{a} \to \mathbf{b}]$? Yuval Peres: *Probabilistically the answer is not obvious. In the language of electrical networks, the question is answered by*

Rayleigh's Monotonicity Law. [Figure 29 and Figure 30 on p.51-52 of Doyle and Snell] If $\{r_e\}$ and $\{r'_e\}$ are two different sets of resistances and if $r_e \leq r'_e$ for all e, then the resistance $\mathcal{R}(\mathbf{a} \leftrightarrow \mathbf{b}) \leq \mathcal{R}'(\mathbf{a} \leftrightarrow \mathbf{b})$, where $\mathcal{R}'$ corresponds to $\{r'_e\}_e$ set.

Thus decreasing the resistance of an existing edge increases $\mathbf{Prob}[\mathbf{a} \to \mathbf{b}]$. Therefore decreasing the resistance of an edge with conductivity zero from infinity to a finite number increases $\mathbf{Prob}[\mathbf{a} \to \mathbf{b}]$.

The proof of the Rayleigh's Monotonicity Law follows form Thompson's Principle:

$$\inf_\theta \sum_e r_e[\theta(e)]^2 \leq \inf_\theta \sum_e r'_e[\theta(e)]^2.$$

1.2.2 Infinite electrical networks.

For a reversible Markov chain on an *infinite* graph G containing vertex $\mathbf{a}$, the same weighted random walk representation with conductance values (and reciprocal resistance values) stands: [From Y.Peres notes.] Let $\{G_n\}$ be a collection of finite connected subgraphs containing $\mathbf{a}$ and satisfying $\cup_n G_n = G$. If all the vertices in $G \setminus G_n$ are replaced by a single vertex $\mathbf{b}_n$, then we can define

$$\mathcal{R}(\mathbf{a} \leftrightarrow \infty) = \lim_{n\to\infty} \mathcal{R}(\mathbf{a} \leftrightarrow \mathbf{b}_n).$$

Then $\mathcal{C}(\mathbf{a} \leftrightarrow \infty) = \frac{1}{\mathcal{R}(\mathbf{a} \leftrightarrow \infty)}$ and

$$\mathbf{Prob}[\mathbf{a} \to \infty] = \frac{\mathcal{C}(\mathbf{a} \leftrightarrow \infty)}{\pi(\mathbf{a})},$$

where $\pi(\mathbf{a}) = \sum_{x:x\sim\mathbf{a}} c_{ax}$.

Thomson's Principle remains valid for infinite networks as well:

$$\mathcal{R}(\mathbf{a} \leftrightarrow \infty) = \inf\{\mathcal{E}(\theta) \ : \ \theta \text{ is a unit flow from } \mathbf{a} \text{ to } \infty\}.$$

An edge-cutset $\prod$ separating $\mathbf{a}$ from $\mathbf{b}$ is a set of edges of a connected graph such that any path from $\mathbf{a}$ to $\mathbf{b}$ must include some edge in $\prod$.

Nash-Williams Inequality (Rayleigh's Method, first applied by Nash-Williams). [From Y.Peres]

If $\{\prod_n\}$ are disjoint edge-cutsets on a connected graph (network) G which separate $\mathbf{a}$ from $\mathbf{b}$, then

$$\mathcal{R}(\mathbf{a} \leftrightarrow \mathbf{b}) \geq \sum_n \left(\frac{1}{\sum_{e\in\prod_n} c_e} \right) .$$

The above inequality works for $\mathbf{b} = \infty$ (i.e. $\lim_{n\to\infty} \mathbf{b}_n$), in which case if

$$\sum_n \left(\frac{1}{\sum_{e\in\prod_n} c_e} \right) = \infty ,$$

the weighted random walk is recurrent.

Proof: If θ is a unit flow from $\mathbf{a}$ to $\mathbf{b}$, then for any n,

$$\left(\sum_{e\in\prod_n} c_e \right) \cdot \left(\sum_{e\in\prod_n} r_e[\theta(e)]^2 \right) \geq \left(\sum_{e\in\prod_n} \sqrt{c_e r_e} |\theta(e)| \right)^2 = \left(\sum_{e\in\prod_n} |\theta(e)| \right)^2 \geq \|\theta\|^2 = 1$$

as $\prod_n$ is an edge-cutset. Therefore

$$\mathcal{E}(\theta) = \sum_e r_e[\theta(e)]^2 \geq \sum_n \sum_{e\in\prod_n} r_e[\theta(e)]^2 \geq \sum_n \left(\frac{1}{\sum_{e\in\prod_n} c_e} \right) .$$

The theorem follows from Thomson's Principal.

□

Example. *Simple random walk on $\mathbb{Z}^2$ is recurrent.* Take $c_e = 1$ for each edge e of $G = \mathbb{Z}^2$ and consider the edge-cutsets $\prod_n$ consisting of of edges joining vertices in $\partial\Box_n$ to vertices in $\partial\Box_{n+1}$, where $\Box_n = [-n, n]^2$. Each $\prod_n$ has $4(2n+1)$ edges in it. Thus

$$\sum_{e\in\prod_n} c_e = 4(2n+1)$$

and, by Nash-Williams,

$$\mathcal{R}(\mathbf{a} \leftrightarrow \infty) \geq \sum_n \frac{1}{4(2n+1)} = \infty.$$

Thus the simple random walk on $\mathbb{Z}^2$ is recurrent.

Example. *Simple random walk on $\mathbb{Z}^3$ is transient.* [Almost entirely from Y.Peres.] Take $c_e = 1$ for each edge e of $G = \mathbb{Z}^3$. To each directed edge $\overrightarrow{e}$ in the lattice $\mathbb{Z}^3$, attach an orthogonal unit square $\Box_e$ intersecting $\overrightarrow{e}$ at its midpoint m_e. Define $\theta(\overrightarrow{e})$ to be the area of the radial projection of $\Box_e$ onto the sphere $\partial B(0, \frac{1}{4})$, taken with a positive sign if the dot product $\overrightarrow{e} \cdot m_e > 0$, and with a negative sign if $\overrightarrow{e} \cdot m_e < 0$. By considering a unit cube centered at each lattice point and projecting it to $\partial B(0, \frac{1}{4})$, we can easily verify that θ satisfies the Kirchhoff's node law at all vertices except the origin. Hence θ is a flow from $\mathbf{a} = 0$ to $\mathbf{b} = \infty$ in $\mathbb{Z}^3$. It is easy to bound its energy:

$$\mathcal{E}(\theta) \leq \sum_n C_1 n^2 \left(\frac{C_2}{n^2}\right)^2 < \infty.$$

Here the number of edges $\overrightarrow{e}$ touching the boundary of $\Box_n = [-n, n]^3$ is bounded by $C_1 n^2$, where for each such $\overrightarrow{e}$, $|\theta(\overrightarrow{e})| \leq \frac{C_2}{n^2}$. Thus $\mathcal{R}(0 \leftrightarrow \infty) < \infty$ and $\mathcal{C}(0 \leftrightarrow \infty) > 0$. Therefore,

$$\mathbf{Prob}[0 \to \infty] = \frac{\mathcal{C}(0 \leftrightarrow \infty)}{\pi(0)} = \frac{\mathcal{C}(0 \leftrightarrow \infty)}{6} > 0.$$

1.3 Martingales and Lyapunov functions

Bremaud Ch.5

Definition. A homogeneous Markov chain $\{X_n\}$ such that $E[|X_n|] < \infty$ (or $X_n \geq 0$) for all $n \geq 0$ is a *martingale* if

$$E[X_{n+1} \mid X_n] = X_n \ .$$

Definition. Given a Markov chain $\{X_n\}$, let $\{Y_n\}$ be a real-valued process such that for each $n \geq 0$,

- Y_n is $\mathcal{F}_n$-measurable, i.e. Y_n is a function of $X_0, \ldots, X_n$;
- $E[|Y_n|] < \infty$ or $Y_n \geq 0$

is called a *martingale* with respect to $\{X_n\}$ if

$$E[Y_{n+1} \mid \mathcal{F}_n] = Y_n \ .$$

If $E[Y_{n+1} \mid \mathcal{F}_n] \leq Y_n$ it is called a *supermartingale*, and if $E[Y_{n+1} \mid \mathcal{F}_n] \geq Y_n$ it is called a *submartingale*.

1.3.1 Martingales and harmonic functions

If $\{X_n\}$ is a homogeneous Markov chain (HMC), and if $h(\cdot)$ is a harmonic function with respect to the transition probabilities $\{p(i, j)\}$, i.e. if h satisfies the averaging property

$$\sum_y p(x, y) h(y) = h(x),$$

then $h(X_n)$ is a martingale. Indeed $E[h(X_{n+1}) \mid X_n = x] = \sum_y p(x,y)h(y) = h(x)$, so

$$E[h(X_{n+1}) \mid \mathcal{F}_n] = h(X_n).$$

The equation $\sum_y p(x,y)h(y) = h(x)$ is often written in operator form as

$$Ph = h \ .$$

1.3.2 Stopping theorem

Bremaud p.185

Theorem 2. *Suppose $\{M_n\}$ is a martingale with respect to $\{X_n\}$, and T is a stopping time with respect to X_n. If either T is bounded or* $\mathbf{Prob}[T < \infty] = 1$ *and there is $K > 0$ such that $|M_n| \leq K$ when $n < T$, then*

$$E[M_T] = E[M_0].$$

Proof: Consider $Y_n = M_{T\wedge n}$, then Y_n is a martingale w.r.t. $\{X_n\}$ and therefore

$$E[M_{T\wedge n}] = E[Y_n] = E[Y_0] = E[M_0].$$

Thus

$$|E[M_T] - E[M_0]| = |E[M_T] - E[M_{T\wedge n}]| \leq 2K\mathbf{Prob}[T > n] \to 0 \quad \text{as} \quad n \to \infty.$$

□

Example. *Gambler's ruin.*

Example. *Birth-and-death chain*

1.3.3 Lyapunov functions

Bremaud Ch.5 and Durrett

Harmonic functions can be generalized to Lyapunov functions, the latter in reversible case

Suppose $\{X_n\}$ is an irreducible Markov chain on state space $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$, and suppose there is a non-negative function ϕ such that

$$\lim_{x\to\infty} \phi(x) = \infty, \quad \text{and} \quad E[\phi(X_{n+1}) \mid X_n = x] \leq \phi(x) \quad \text{when } x \geq K,$$

for some $K > 0$. Then we can show that $\{X_n\}$ is recurrent.

Indeed we can show that the Markov chain visits $\{0, 1, \ldots, K\}$ infinitely often. Let τ_K be the time of hitting set $\{0, 1, \ldots, K\}$. We fix $M > K$, and let τ_M be the time of hitting set $\{M, M+1, \ldots\}$. Then $\tau := \min\{\tau_K, \tau_M\}$ is a stopping time.

Without loss of generality we let $\phi \equiv 0$ on $\{0, 1, \ldots, K-1\}$. Now, $\phi(X_n)$ is a bounded supermartingale on $\{K, K+1, \ldots, M\}$. Thus

$$E[\phi(X_\tau) \mid X_0 = x] \le \phi(x).$$

Therefore

$$(1 - \mathbf{Prob}[\tau_K < \tau_M]) \cdot E[\phi(X_{\tau_M}) \mid X_0 = x,\ \tau_M < \tau_K] = \phi(x).$$

Thus, for any $x \in \mathbb{Z}_+$,

$$\mathbf{Prob}[\tau_K < \tau_M] = 1 - \frac{\phi(x)}{E[\phi(X_{\tau_M}) \mid X_0 = x,\ \tau_M < \tau_K]} \to 1 \quad \text{as} \quad M \to +\infty$$

as $E[\phi(X_{\tau_M}) \mid X_0 = x,\ \tau_M < \tau_K] \ge \min\{\phi(y) \ : \ y = M, M+1, \ldots\} \to \infty$.

Hence $\mathbf{Prob}[\tau_K < \infty] = \lim_{M \to +\infty} \mathbf{Prob}[\tau_K < \tau_M] = 1$ and $\mathbf{Prob}[T_0 < \infty] = 1$, i.e. the chain is *recurrent.*

Such $\phi(\cdot)$ is called a **Lyapunov** function by analogy with Lyapunov functions used to show stability of differential equations. In general, the Lyapunov functions can be used just as well in higher dimensions. A similar condition for $\mathbb{Z}^d$ is

$$\lim_{\|x\| \to \infty} \phi(x) = \infty, \quad \text{and} \quad E[\phi(X_{n+1}) \mid X_n = x] \le \phi(x) \quad \text{when } \|x\| \ge K,$$

for some $K > 0$. The proof of recurrence is as above.

1.3.4 Example: 1-D Random Walk in Random Environment.

Consider a sequence of i.i.d. $(0,1)$-valued random variables $p_0, p_1, p_2, \ldots$ and a Markov chain (random walk) $\{X_n\}$ on $\mathbb{Z}_+ = \{0, 1, \ldots\}$ with the following transition probabilities: $p(0,1) = p_0 = 1 - p(0,0)$,

$$p(j, j+1) = p_j \quad \text{and } p(j, j-1) = q_j = 1 - p_j \quad \text{for all} \quad j = 1, 2, \ldots.$$

Here $\{p_j\}$ is called random environment, and $\{X_n\}$ is a **random walk in random environment** (RWRE). Let

$$\rho_j := \frac{1 - p_j}{p_j}.$$

Theorem 3. *(F.Salomon 1975) The random walk in random environment is*

- *recurrent with probability one if* $\mu = E[\log \rho_j] \ge 0$
- *transient with probability one if* $\mu = E[\log \rho_j] < 0$

Proof via martingales / Lyapunov functions: We let

$$\phi(x) = 1 + \rho_1 + \rho_1\rho_2 + \rho_1\rho_2\rho_3 + \rho_1\rho_2\rho_3\rho_4 + \cdots + \rho_1\rho_2\ldots\rho_{x-1}.$$

Then for each $x \geq 1$, $\phi(x+1) - \phi(x) = \rho_1\rho_2\ldots\rho_{x-1}\rho_x = (\phi(x) - \phi(x-1))\rho_x$ and therefore $p_x(\phi(x+1) - \phi(x)) = (1-p_x)(\phi(x) - \phi(x-1))$. So

$$\phi(x) = p_x\phi(x+1) + (1-p_x)\phi(x-1).$$

Hence, conditioned on the environment $\{p_j\}$, $\phi(X_n)$ is a martingale on $\{1, 2, \ldots\}$:

$$E[\phi(X_{n+1}) \mid X_n,\ \{p_j\}] = \phi(X_n).$$

Thus for $M > 0$ and $x \in \{0, 1, 2, \ldots, M\}$,

$$\mathbf{Prob}[T_M < T_0] = \frac{\phi(x) - \phi(0)}{\phi(M) - \phi(0)}.$$

In order to show recurrence we need to show $\lim_{x\to\infty}\phi(x) = \infty$. Now, by the strong law of large numbers,

$$\mathbf{Prob}\left[\frac{\log\rho_1 + \log\rho_2 + \cdots + \log\rho_n}{n} \to \mu\right] = 1.$$

If $\mu > 0$,

$$\rho_1\rho_2\ldots\rho_x = \exp\left\{\frac{\log\rho_1 + \log\rho_2 + \cdots + \log\rho_x}{x}\cdot x\right\} \geq \left(1 + \frac{\mu-1}{2}\right)^x > 1$$

for x large enough and $\lim_{x\to\infty}\phi(x) = \infty$ with probability one, thus proving recurrence.

If $\mu < 0$,

$$\rho_1\rho_2\ldots\rho_x = \exp\left\{\frac{\log\rho_1 + \log\rho_2 + \cdots + \log\rho_x}{x}\cdot x\right\} \leq \left(1 - \frac{1-\mu}{2}\right)^x$$

for large x and $\lim_{x\to\infty}\phi(x) < \infty$ each time, thus proving transience.

Now, if $\mu = 0$, one can show that the random walk process

$$Y_n = \log\rho_1 + \log\rho_2 + \cdots + \log\rho_n,$$

where steps $\{\log\rho_j\}$ are mean zero i.i.d. random variables, returns to $\mathbb{R}_+$ infinitely often: Y_n restricted to $[0, -K]$ $(K > 0)$ is a bounded martingale, and the probability of getting to $\mathbb{R}_+$ before $(-\infty, -K]$ increases to one as $K \to +\infty$. Thus

$$\rho_1\rho_2\ldots\rho_x \geq 1$$

infinitely often and $\lim_{x\to\infty}\phi(x) = \infty$ implying recurrence. *Q.E.D.*

Proof via electrical networks: The edge from 0 to itself has conductivity c_0, edge $[0,1]$ has conductivity c_1, edge $[1,2]$ has conductivity c_2, $[2,3]$ has conductivity c_2 and so on.

Now, $\pi(0) = c_0 + c_1$, $\pi(1) = c_1 + c_2$, $\pi(2) = c_2 + c_3$ etc. So

$$c_1 = \pi(0)p_0,\ c_2 = \pi(1)p_1,\ c_3 = \pi(2)p_2,\ \ldots,$$

where site conductivity $\pi(\cdot)$ satisfies detailed balance conditions and therefore $\pi(1) = \frac{p_0}{(1-p_1)}\pi(0)$,

$$\pi(2) = \frac{p_1}{(1-p_2)}\pi(1) = \frac{p_0 p_1}{(1-p_1)(1-p_2)}\pi(0),$$

$$\pi(3) = \frac{p_2}{(1-p_3)}\pi(2) = \frac{p_0 p_1 p_2}{(1-p_1)(1-p_2)(1-p_3)}\pi(0),$$

$$\cdots$$

$$\pi(j) = \frac{p_{j-1}}{(1-p_j)}\pi(j-1) = \frac{p_0 p_1 \ldots p_{j-1}}{(1-p_1)(1-p_2)\ldots(1-p_j)}\pi(0),$$

and so on. Thus

$$c_2 = p_1\pi(1) = \frac{c_1}{\rho_1},\ c_3 = p_2\pi(2) = \frac{c_1}{\rho_1\rho_2},\ c_4 = p_3\pi(3) = \frac{c_1}{\rho_1\rho_2\rho_3}, \ldots$$

and

$$\mathcal{R}(0 \leftrightarrow \infty) = \frac{1}{c_1}\left(1 + \rho_1 + \rho_1\rho_2 + \rho_1\rho_2\rho_3 + \ldots\right).$$

The rest of the proof is as above. *Q.E.D.*

1.3.5 Martingale convergence theorem

Bremaud, p.185

Theorem 4. *Let $\{Y_n\}$ be either a nonnegative supermartingale, or a bounded submartingale, with respect to $\{X_n\}$. Then, with probability one, $\lim_{n\to\infty} Y_n$ exists and is finite.*

The proof is done via the upcrossing inequality.

1.4 Homework #1

Problem 1. *2-D RWRE* Consider two sequences of i.i.d. $(0,1)$-valued random variables $p_1, p_2, p_3, \ldots$ and $\bar{p}_1, \bar{p}_2, \bar{p}_3, \ldots$ and a Markov chain (random walk) $\{X_n\}$ on $\mathbb{Z}_+^2 = \{0, 1, \ldots\}^2$ with the following transition probabilities:

$$p((i,j),(i+1,j)) = \frac{p_i}{2}, \qquad p((i,j),(i-1,j)) = \frac{1-p_i}{2},$$

$$p((i,j),(i,j+1)) = \frac{\bar{p}_j}{2}, \quad \text{and} \quad p((i,j),(i,j-1)) = \frac{1-\bar{p}_j}{2} \qquad \text{for all} \quad i,j \in \mathbb{Z}_+^2,$$

where $p_0 = \bar{p}_0 = 1$. Suppose $\mu = E\left[\log\left(\frac{1-p_1}{p_1}\right)\right] > 0$ and $\bar{\mu} = E\left[\log\left(\frac{1-\bar{p}_1}{\bar{p}_1}\right)\right] > 0$. Show that $\{X_n\}$ is recurrent.

Problem 2. Suppose $\xi_1, \xi_2, \ldots$ are i.i.d. with mean $E[\xi_i] = 0$. Let $S_n = \xi_1 + \xi_2 + \cdots + \xi_n$ be a random walk. Show that $\{S_n\}$ visits $\mathbb{R}_+ = \{x \ : \ x \geq 0\}$ infinitely often. Observe that ξ_i could have unbounded higher moments, e.g. $E[\xi_i^2] = \infty$.

Chapter 2

Long Run Behavior of Stochastic Processes.

Bremaud Ch.6

2.1 Coupling method, convergence rates via coupling.

Bremaud, p.129; Convergence to steady state: Liggett, p.65; Bremaud, p.128-131

Theorem 5. *Let $\{p(i,j)\}$ be the transition probabilities for an irreducible and aperiodic Markov chain on a finite set S, i.e. $\exists m$ s.t. $\min_{i,j\in S} p_m(i,j) = \varepsilon > 0$. Then the limit*

$$\pi(j) = \lim_{n\to\infty} p_n(i,j)$$

exists for all $i, j \in S$ and is independent of i, where π is the unique stationary distribution.

Coupling proof: [Liggett, p.65] Let (X_n, Y_n) be a Markov chain on $S \times S$ with transition probabilities

$$\mathbf{Prob}[(X_{n+1}, Y_{n+1}) = (j_1, j_2) \mid (X_n, Y_n) = (i_1, i_2)] = \begin{cases} p(i_1, j_1)p(i_2, j_2) & \text{if } i_1 \neq i_2, \\ p(i_1, j_1) & \text{if } i_1 = i_2 \text{ and } j_1 = j_2, \\ 0 & \text{if } i_1 = i_2 \text{ but } j_1 \neq j_2. \end{cases}$$

So until the coupling time $T = \min\{n \geq 0 \mid X_n = Y_n\}$ the two processes $\{X_n\}$ and $\{Y_n\}$ move as two independent Markov chains with transition probabilities $\{p(i,j)\}$. After T, they move as one Markov chain

$$X_T = Y_T, \quad X_{T+1} = Y_{T+1}, \quad X_{T+2} = Y_{T+2}, \quad X_{T+3} = Y_{T+3}, \quad \ldots$$

with transition probabilities $\{p(i,j)\}$. Now

$$\begin{aligned} \mathbf{Prob}[T \leq m \mid (X_0, Y_0) = (i_x, i_y)] &= \sum_{j\in S} \mathbf{Prob}[X_m = Y_m = j \mid (X_0, Y_0) = (i_x, i_y)] \\ &\geq \sum_{j\in S} p_m(i_x, j) p_m(i_y, j) \geq \varepsilon \end{aligned}$$

for any $i_x, i_y \in S$. Thus $\mathbf{Prob}[T < \infty] = 1$ and

$$\mathbf{Prob}[X_n \neq Y_n \mid (X_0, Y_0) = (i_x, i_y)] \to 0$$

as $n \to \infty$. Hence, for any $i_x, i_y \in S$,.

$$|p_n(i_x, j) - p_n(i_y, j)| = |\mathbf{Prob}[X_n = j \mid (X_0, Y_0) = (i_x, i_y)] - \mathbf{Prob}[Y_n = j \mid (X_0, Y_0) = (i_x, i_y)]|$$

$$= |\mathbf{Prob}[X_n = j, T > n \mid (X_0, Y_0) = (i_x, i_y)] - \mathbf{Prob}[Y_n = j, T > n \mid (X_0, Y_0) = (i_x, i_y)]|$$

$$\leq \mathbf{Prob}[X_n \neq Y_n \mid (X_0, Y_0) = (i_x, i_y)] \to 0$$

□

Here we will quote [R.Durrett, "Probability: Theory and Examples."]:
"**Example. A coupling card trick.** The following demonstration used by E.B.Dynkin in his probability class is a variation of a card trick that appeared in Scientific American. The instructor asks a student to write 100 random digits from 0 to 9 on the blackboard. Another student chooses one of the first 10 numbers and does not tell the instructor. If that digit is 7 say she counts 7 places along the list, notes the digit at that location, and continues the process. If the digit is 0 she counts 10. A possible sequence is underlined on the list below:

$$3\ 4\ \underline{7}\ 8\ 2\ 3\ 7\ 5\ 6\ \underline{1}\ \underline{6}\ 4\ 6\ 5\ 7\ 8\ \underline{3}\ 1\ 5\ \underline{3}\ 0\ 7\ \underline{9}\ 2\ 3\ \ldots$$

The trick is that, without knowing the student's first digit, the instructor can point to her final stopping position. To this end, he picks the first digit, and forms his own sequence in the same manner as the student and announces his stopping position. He makes an error if the coupling time is larger than 100. Numerical computation done by one of Dynkin's graduate students show that the probability of error is approximately .026"

2.2 Second largest eigenvalue.

Linear algebra is central to Markov chains. Let us recall a few facts: Suppose A is an $r \times r$ matrix with all non-negative entries and r distinct eigenvalues $\lambda_1, \ldots, \lambda_r$, and $u_1, \ldots, u_r$ and $v_1, \ldots, v_r$ are respectively left and right eigenvectors (1 row, r columns), i.e.

$$u_i A = \lambda_i u_i \quad \text{and} \quad A v_i^T = \lambda_i v_i^T \quad i = 1, 2, \ldots, r.$$

Then $\lambda_i u_i v_j^T = u_i A v_j^T = \lambda_j u_i v_j^T$ and therefore $u_i \cdot v_j = u_i v_j^T = 0$ if $i \neq j$. One can scale the eigenvectors so that $u_i \cdot v_i = u_i v_i^T = 1$ for all i. So A is diagonalizable, i.e.

$$\Lambda = U A V^T,$$

where $u_1, \ldots, u_r$ are rows of U, $v_1, \ldots, v_r$ are rows of V and Λ is the diagonal matrix of eigenvalues $\lambda_1, \ldots, \lambda_r$. Here $U^{-1} = V^T$.

Thus we obtain spectral decomposition

$$A^n = V^T \Lambda U = \sum_{i=1}^{r} \lambda_i^n v_i^T u_i$$

Consider a finite Markov chain on state space S of cardinality $|S| = r$ with transition probabilities matrix $P = \{p(i,j)\}$ and unique stationary distribution π. Let

$$1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_r > -1$$

be the eigenvalues of P. Since P is a nonnegative matrix whose rows add up to one, if there was an eigenvalue λ with $|\lambda| > 1$, then there is a nonnegative vector u with $\|u\|_{l^1} = 1$ (i.e. its coordinates add up to one) such that $\|uP\|_{l^1} > 1$, but $1 = \|u\|_{l^1} = \|uP\|_{l^1}$.

Now $\lambda_1 = 1$, $u_1 = \pi$ and $v_1 = \mathbf{1} = (1, 1, \ldots, 1)$ as $\pi P = \pi$ and the rows of P add up to one.

Example. *Two state Markov chain.* See Bremaud, p.196.

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

The eigenvalues are $\lambda_1 = 1 > \lambda_2 = 1 - \alpha - \beta > -1$ with

$$u_1 = \pi = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta} \right), \qquad v_1 = \mathbf{1} = (1,1)$$

and

$$u_2 = (1, -1), \qquad v_2 = \left(\frac{\alpha}{\alpha+\beta}, \frac{-\beta}{\alpha+\beta} \right).$$

Thus

$$P^n = \sum_{i=1}^{2} \lambda_i^n v_i^T u_i = \frac{1}{\alpha+\beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1-\alpha-\beta)^n}{\alpha+\beta} \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}$$

and we obtain the result of the limit theorem that we already proved with coupling.
Thus

$$P^n \to \Pi = \frac{1}{\alpha+\beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 \\ \pi_1 & \pi_2 \end{pmatrix} \qquad \text{where } \pi = (\pi_1, \pi_2).$$

Here it is important that we also know the rate of convergence

$$P^n - \Pi = \frac{(1-\alpha-\beta)^n}{\alpha+\beta} \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}.$$

Convergence to steady state via Perron-Frobeniuous theorem. A finite state Markov chain is irreducible and aperiodic if there is $m > 0$ such that P^m has all positive entries, we say $P^m > 0$. A nonnegative matrix A that satisfies this condition ($A^m > 0$ for some m) is said to be *primitive*.

Theorem 6. (Perron-Frobenius) *If A is a nonnegative primitive matrix, then there is an eigenvector λ of multiplicity one such that $\lambda_1 > |\lambda_j|$ for all other eigenvalues λ_j. Moreover the corresponding left and right eigenvectors u_1 and v_1 can be chosen positive and such that $u_1 v_1^T = 1$.*

Now if $\lambda_{(2)}$ is the second largest in absolute value eigenvector of P, and m_2 is its algebraic multiplicity, then

$$P^n = \lambda_1^n v_1^T u_1 + O(n^{m_2-1}|\lambda_{(2)}|^n) = \Pi + O(n^{m_2-1}|\lambda_{(2)}|^n)$$

thus obtaining the rates via Perron-Frobenius theorem.

Remark. If the finite Markov chain generated by P irreducible, but not aperiodic. Then if we define a new transition probability matrix $\widehat{P} = \frac{1}{2}I + \frac{1}{2}P$, i.e. each time we toss a coin and either do nothing with probability $\frac{1}{2}$, or take a step according to P. Then $\widehat{P}$ is irreducible and aperiodic, and thus has the unique stationary distribution π. But $\pi P = \pi$ if and only if $\pi \widehat{P} = \pi$. Thus P has a *unique* stationary distribution. One can show that all the eigenvalues of $\widehat{P}$ are nonnegative.

2.2.1 Spectral gap and spectral theorem.

Bremaud, p.195-221

There are two ways to label the eigenvalues: We can label the eigenvalues in decreasing order

$$1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_r > -1,$$

or we can order them in decreasing order with respect to absolute values

$$1 = \lambda_{(1)} > |\lambda_{(2)}| \geq |\lambda_{(3)}| \geq \cdots \geq |\lambda_{(r)}| \geq 0.$$

Value $1 - \lambda_2$ is called the **spectral gap**, and value $1 - |\lambda_{(2)}|$ is called the **absolute spectral gap**.

Now if we slow down the Makov chain by skipping the move each time with probability $\frac{1}{2}$, i.e. considering $\widehat{P} = \frac{1}{2}I + \frac{1}{2}P$. Then for the $\widehat{P}$ chain, $\lambda_2 = \lambda_{(2)}$ and the spectral gap is the same as absolute spectral gap. Observe that in the long run (n is large) $\widehat{P}$ is only twice slower than P.

Therefore for the rest of the section we can **assume** without loss of generality that

$$\lambda_2 = \lambda_{(2)}$$

as well as that the Markov chain is not only irreducible, but also aperiodic. We also assume that it is **reversible**.

Example. Relaxation time is the reciprocal of the spectral gap:

$$\tau_{rlx} = \frac{1}{1 - \lambda_2}.$$

Why relaxation time? Suppose $\lambda_2 = \lambda_{(2)}$ has multiplicity $m_2 = 1$. Then after $n = K \cdot \tau_{rlx}$ iterations the distribution tail

$$P^n - \Pi = O(\lambda_2^n),$$

where

$$\lambda_2^n = e^{\frac{\log \lambda_2}{1-\lambda_2}K} = e^{-(1+\frac{(1-\lambda_2)}{2}+\frac{(1-\lambda_2)^2}{3}+\dots)K} \leq e^{-K}.$$

Now, we assumed reversibility

$$\pi(i)p(i,j) = \pi(j)p(j,i).$$

We can define the inner product with respect to π,

$$< x, y >_\pi := \sum_{i,j \in S} x(i)y(i)\pi(i)$$

and the corresponding $l^2(\pi)$ norm

$$\|x\|_\pi = \sqrt{< x, x >_\pi}.$$

Observe that $< x, \mathbf{1} >_\pi = E_\pi[x]$ - the mean of x w.r.t. probability distribution π and $\|x\|_\pi = E_\pi[x^2]$ is the second moment.

In general, if μ is a measure on S, we let $l^2(\mu)$ denote the r-dimensional space $\mathbb{R}^r$ with inner product

$$< x, y >_\mu = \sum_{i,j \in S} x(i)y(i)\mu(i)$$

Theorem 7. *In the reversible case (which we assumed), P is self-adjoined in $l^2(\pi)$, i.e.*

$$< xP^T, y >_\pi = < x, yP^T >_\pi$$

for all $x, y \in l^2(\pi)$.

Proof:

$$< xP^T, y >_\pi = \sum_{i,j} \pi(i)p(i,j)x(j)y(i) = \sum_{i,j} \pi(j)p(j,i)x(j)y(i) = < x, yP^T >_\pi$$

□

Converse is also true: if P is self-adjoint then the chain is reversible. This follows from taking $x = \vec{e}_i$ and $y = \vec{e}_j$ in $\mathbb{R}^r$ and substituting into $< xP^T, y >_\pi = < x, yP^T >_\pi$.

We let

$$D = \begin{pmatrix} \pi(1) & 0 & 0 & 0 \\ 0 & \pi(2) & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \pi(r) \end{pmatrix} \qquad \text{and} \qquad P^* = D^{\frac{1}{2}} P D^{-\frac{1}{2}}.$$

Here the eigenvalues of P^* are the same $\lambda_1, \dots, \lambda_r$ as those for P. Moreover P^* is **symmetric** (i.e. $(P^*)^T = P^*$) as P is reversible w.r.t. π and

$$p^*(i,j) = \sqrt{\frac{\pi(i)}{\pi(j)}}p(i,j) = \sqrt{p(i,j)p(j,i)} = \sqrt{\frac{\pi(j)}{\pi(i)}}p(j,i) = p^*(j,i).$$

Thus P^* has an orthonormal set of left eigenvalues $w_1, \dots, w_r$, where by symmetry $w_1^T, \dots, w_r^T$ are corresponding right eigenvalues. Then $u_1, \dots, u_r$ and $v_1, \dots, v_r$ defined as

$$u_i = w_i D^{\frac{1}{2}} \qquad \text{and} \qquad v_i^T = D^{-\frac{1}{2}} w_i^T$$

are respectively the left and the right eigenvalues of P. Observe that

$$u_i = v_i D.$$

Now, $xDy^T = < x, y >_\pi$ and therefore

$$< v_i, v_j >_\pi = \delta_{ij} \qquad \text{and} \qquad < u_i, u_j >_{\frac{1}{\pi}} = \delta_{ij}$$

as $< v_i, v_j >_\pi = v_i D v_j^T = (w_i D^{-\frac{1}{2}}) D (D^{-\frac{1}{2}} w_j^T) = w_i w_j^T = \delta_{ij}$, i.e. $u_1, \dots, u_r$ are orthonormal w.r.t. $< \cdot, \cdot >_{\frac{1}{\pi}}$ and $v_1, \dots, v_r$ are orthonormal w.r.t. $< \cdot, \cdot >_\pi$. Hence

$$x = \sum_{j=1}^r < x, u_j >_{\frac{1}{\pi}} u_j \qquad \text{and} \qquad x = \sum_{j=1}^r < x, v_j >_\pi v_j$$

and

$$xP^n = \sum_{j=1}^r \lambda_j^n < x, u_j >_{\frac{1}{\pi}} u_j \qquad \text{and} \qquad P^n x^T = \sum_{j=1}^r \lambda_j^n < x, v_j >_\pi v_j^T.$$

Recall that $u_1 = \pi$ and $v_1 = \mathbf{1}$, and therefore $w_1 = (\sqrt{\pi(1)}, \dots, \sqrt{\pi(r)})$.

Definition. Dirichlet form

$$\mathcal{E}_\pi(x,x) = < x(I-P)^T, x >_\pi = < x(I - P^T), x >_\pi \quad .$$

We will adapt the following notation: $\mathcal{E}(x,x) = \mathcal{E}_\pi(x,x)$.

Theorem 8. $\mathcal{E}(x,x) = \frac{1}{2} \sum_{i,j \in S} \pi(i) p(i,j) [x(j) - x(i)]^2$

Proof:

$$\begin{aligned}
< x(I-P)^T, x >_\pi &= \sum_i \left(x(i) - \sum_j p(i,j)x(j) \right) x(i)\pi(i) \\
&= \sum_i \left(\sum_j p(i,j)[x(i)-x(j)] \right) x(i)\pi(i) \\
&= \sum_{i,j} [x(i)-x(j)]x(i)\pi(i)p(i,j) \\
&= \sum_{i,j} [x(i)-x(j)]x(j)\pi(j)p(j,i) \\
&= \frac{1}{2}\sum_{i,j} [x(i)-x(j)]x(i)\pi(i)p(i,j) + \frac{1}{2}\sum_{i,j} [x(i)-x(j)]x(j)\pi(j)p(j,i) \\
&= \frac{1}{2}\sum_{i,j\in S} \pi(i)p(i,j)[x(j)-x(i)]^2
\end{aligned}$$

□

Recall that in the electrical networks representation of reversible chains,

$$\mathcal{E}(I) = \frac{1}{2}\sum_{i,j} \pi(i)p(i,j)\left(V(j)-V(i)\right)^2 \ .$$

Also, in general, the same proof works in showing

$$\mathcal{E}(x,y) := < x(I-P^T), y >_\pi = \frac{1}{2}\sum_{i,j\in S} \pi(i)p(i,j)(x(j)-x(i))(y(j)-y(i)) \ .$$

Theorem 9. *(Rayleigh's Spectral Theorem for the second largest eigenvalue.)*

$$1-\lambda_2 = \inf\left\{ \frac{\mathcal{E}(x,x)}{\|x\|_\pi^2} \ : \ < x, \mathbf{1} >_\pi = \mathbf{0} \right\}$$

In general, Rayleigh's Theorem asserts

$$1-\lambda_k = \inf\left\{ \frac{\mathcal{E}(x,x)}{\|x\|_\pi^2} \ : \ < x, v_i >_\pi = 0, \ 1 \le i < k \right\}.$$

Proof:

$$(I-P)x^T = \sum_{j=1}^r < x, v_j >_\pi v_j^T - \sum_{j=1}^r \lambda_j < x, v_j >_\pi v_j^T = \sum_{j=1}^r (1-\lambda_j) < x, v_j >_\pi v_j^T$$

and

$$x(I-P)^T = \sum_{j=1}^{r}(1-\lambda_j) < x, v_j >_\pi v_j \ .$$

Thus

$$\mathcal{E}_\pi(x,x) =< x(I-P)^T, x >_\pi = \sum_{j=1}^{r}(1-\lambda_j) < x, v_j >_\pi^2 \ .$$

Now we compare $\mathcal{E}_\pi(x,x)$ to

$$\|x\|_\pi^2 =< x, x >_\pi = \left\langle x, \ \sum_{j=1}^{r} < x, v_j >_\pi v_j \right\rangle_\pi = \sum_{j=1}^{r} < x, v_j >_\pi^2$$

when $< x, v_1 >_\pi =< x, \mathbf{1} >_\pi = 0$. There

$$\frac{\mathcal{E}(x,x)}{\|x\|_\pi^2} = \frac{\sum_{j=2}^{r}(1-\lambda_j) < x, v_j >_\pi^2}{\sum_{j=2}^{r} < x, v_j >_\pi^2} \geq (1-\lambda_2)$$

as $(1-\lambda_2) \ \leq \ (1-\lambda_3) \ \leq \ (1-\lambda_4) \ \leq \ \ldots \ \leq \ (1-\lambda_r)$, where

$$\inf\left\{\frac{\mathcal{E}(x,x)}{\|x\|_\pi^2} \ : \ < x, \mathbf{1} >_\pi = \mathbf{0}\right\} = 1-\lambda_2$$

is attained at $x = v_2$. □

Thus any $y = (y(1), \ldots, y(r))$ such that $y(1) + \cdots + y(r) = 0$ provides an upper bound on spectral gap

$$(1-\lambda_2) \ \leq \ \frac{\mathcal{E}(y,y)}{\|y\|_\pi^2}$$

and a lower bound on relaxation time

$$\tau_{rlx} \ \geq \ \frac{\|y\|_\pi^2}{\mathcal{E}(y,y)} \ .$$

2.2.2 Relaxation times.

Bremaud, p.212; [Diaconis and Strook, 1991]

The reversibility and assumptions section are imposed. Let us return the random walk on weighted graph construction of electrical networks. So,

$$c_{ij} = \pi(i)p(i,j) \ .$$

For each pair of states $i \neq j$, we randomly select exactly one edge self-avoiding (i.e. no edge is used more than once) path

$$\rho_{ij} = \{i \to i_1 \to i_2 \to \cdots \to i_m \to j\}$$

of positive probability $p(i, i_1)p(i_1, i_2) \dots p(i_m, j) > 0$. Then the resistance of the path is

$$\mathcal{R}(\rho_{i,j}) = r_{ii_1} + r_{i_1 i_2} + \cdots + r_{i_m j} = \frac{1}{\pi(i)p(i, i_1)} + \frac{1}{\pi(i_1)p(i_1, i_2)} + \cdots + \frac{1}{\pi(i_m)p(i_m, j)} \ .$$

Each pair $i \neq j$ has a unique such path in the collection $\mathcal{P}$. We define the **Poincaré coefficient** associated with $\mathcal{P}$

$$\kappa = \kappa(\mathcal{P}) = \max_e \sum_{i,j:\ e \in \rho_{i,j}} \pi(i)\mathcal{R}(\rho_{i,j})\pi(j) \ .$$

Theorem 10. *Given the assumptions of this section,*

$$\lambda_2 \leq 1 - \frac{1}{\kappa} \ .$$

Then $\tau_{rlx} \leq \kappa$.

Proof: For any $x \in \mathbb{R}$ such that $< x, \mathbf{1} >_\pi = 0$,

$$\begin{aligned}
\|x\|_\pi^2 &= < x, x >_\pi = < x, x >_\pi - < x, \mathbf{1} >_\pi^2 \\
&= \frac{1}{2} \sum_{i,j} (x(j) - x(i))^2 \pi(i)\pi(j) \\
&= \frac{1}{2} \sum_{i,j} \left[\sum_{e=[i_-, i_+]\ :\ e \in \rho_{i,j}} \frac{1}{\sqrt{c_e}} \sqrt{c_e} \cdot (x(i_+) - x(i_-)) \right]^2 \pi(i)\pi(j) \\
&\leq \frac{1}{2} \sum_{i,j} \mathcal{R}(\rho_{i,j}) \left[\sum_{e=[i_-, i_+]\ :\ e \in \rho_{i,j}} c_e \cdot (x(i_+) - x(i_-))^2 \right] \pi(i)\pi(j) \quad \text{(Cauchy-Schwarz)} \\
&\leq \frac{1}{2} \sum_{e=[i_-, i_+]} \left(c_e \cdot (x(i_+) - x(i_-))^2 \cdot \left[\sum_{i,j:\ e \in \rho_{i,j}} \pi(i)\mathcal{R}(\rho_{i,j})\pi(j) \right] \right) \\
&\leq \kappa \cdot \mathcal{E}(x, x)
\end{aligned}$$

Thus $\frac{1}{\kappa} \leq \frac{\mathcal{E}(x,x)}{\|x\|_\pi^2}$ for all $x \in \mathbb{R}$ such that $< x, \mathbf{1} >_\pi = 0$. Hence, by the Spectral Theorem,

$$\frac{1}{\kappa} \leq 1 - \lambda_2 \ .$$

□

Example. *Random walk on a graph G.* [Bremaud. p.214] There $\pi(i) = \frac{\deg(i)}{2|\mathcal{E}|}$, where $|\mathcal{E}|$ is the number of edges in G. Also $c_e = \frac{1}{2|\mathcal{E}|}$ for all $e \in \mathcal{E}$ and

$$\kappa(\mathcal{P}) = \max_{e \in \mathcal{E}} \frac{1}{2|\mathcal{E}|} \sum_{i,j:\ e \in \rho_{i,j}} \deg(i) \cdot |\rho_{i,j}| \cdot \deg(j),$$

where $\mathcal{R}(\rho_{i,j}) = 2|\mathcal{E}| \cdot |\rho_{i,j}|$.

We let $d = \max_{i\in G} \deg(i)$, $|\rho| = \max_{i,j\in G} |\rho_{i,j}|$ and define the *bottleneck coefficient*

$$B = \max_{e\in\mathcal{E}}\{\#\text{ of paths } \rho_{i,j} \in \mathcal{P} \text{ s.t. } e \in \rho_{i,j}\}.$$

Then

$$\tau_{rlx} \leq \kappa(\mathcal{P}) \leq \frac{1}{2|\mathcal{E}|}|\rho| d^2 B \ .$$

2.3 Mixing times.

The hitting times are an important tool that we will be using in this section.

Mean hitting time. Consider a simple random walk $\{X_n\}$ on $\mathbb{Z}$. Given two positive integers, a and b, let T_{-a} and T_b be respective hitting times for $-a$ and b. Also let $T = T_{-a} \wedge T_b$. We want to compute $E[T \mid X_0 = 0]$.
Let $g(x) = (b-x)(x+a)$ then g satisfies

$$g(x) = 1 + \sum_y p(x,y) g(y)$$

as $g(x) = 1 + \frac{1}{2}((b-x)-1)((x+a)+1) + \frac{1}{2}((b-x)+1)((x+a)-1)$. Also $g \equiv 0$ on $A = \{-a, b\}$ and T is the hitting time with respect to set A. Thus $M_n = g(X_{T\wedge n}) + (T \wedge n)$ is a martingale w.r.t. $\{X_n\}$. Therefore, by the optional stopping theorem,

$$E[T \mid X_0 = x] = E[M_T \mid X_0 = x] = E[m_0 \mid X_0 = x] = g(x)$$

for all $x \in [-a, b]$. Hence

$$E[T \mid X_0 = 0] = g(0) = ab \ .$$

Thus for the random walk on the ring $\mathbb{Z}/n\mathbb{Z}$,

$$E[T_{ij}] = E[T_j \mid X_0 = i] = (j-i)(n-j+i)$$

for all $i, j \in \{0, 1, \ldots, n-1\}$.

2.3.1 Strong stationary times.

Bremaud, p.219; [Diaconis]

Total variation distance:

$$\|\mu - \nu\|_{TV} := \frac{1}{2}\sum_{x\in S} |\mu(x) - \nu(x)| = \sup_{A\subset S} |\mu(A) - \nu(A)|$$

Definition. Let $\{X_n\}$ be a time-homogeneous Markov chain. A stopping time T is called **strong stationary time** if $\mathbf{Prob}[T < \infty] = 1$, and $X_T \sim \pi$ and is independent of T.

Definition. *Mixing time:* process $\{X_t\}$ with stationary distribution π. If $X_t \sim \nu_t = \nu_0 P_t$,

$$t_{mix} := \inf\left\{t \;:\; \|\nu_t - \pi\|_{TV} \leq \frac{1}{4}, \quad \text{all } \nu_0\right\} .$$

So, if one compares the process with initial condition $X_0 \sim \nu$ with the process with initial conditions $X_0 \sim \pi$,

$$\begin{aligned}
\|\nu P^t - \pi\|_{TV} &= \frac{1}{2}\sum_{x\in S} |\mathbf{Prob}[X_t = x \mid X_0 \sim \nu] - \mathbf{Prob}[X_t = x \mid X_0 \sim \pi]| \\
&= \frac{1}{2}\sum_{x\in S} |\mathbf{Prob}[X_t = x,\ t < T \mid X_0 \sim \nu] - \mathbf{Prob}[X_t = x,\ t < T \mid X_0 \sim \pi]| \\
&\leq \max_{x_0} \mathbf{Prob}[t < T \mid X_0 = x_0] \\
&\leq \frac{\max_{x_0} E[T \mid X_0 = x_0]}{t}
\end{aligned}$$

by Markov inequality. Thus letting $t \geq 4 \cdot \max_{x_0} E[T \mid X_0 = x_0]$ obtain $\|\nu P^t - \pi\|_{TV} \leq \frac{1}{4}$. Thus $t_{mix} \leq 4 \cdot \max_{x_0} E[T \mid X_0 = x_0]$. A sharper estimate can be achieved with the knowledge of higher moments of T.

Example. *Coupon collector.*

Example. Bremaud p.224; originally done in [Diaconis and Fill, 1991] This is the example of "lazy" simple random walk on a one dimensional torus

$$\mathbb{Z}/n\mathbb{Z} = (\mathbb{Z} \mod n) = \{0, 1, \ldots, n-1\}.$$

The random walk is called lazy because of its transition probabilities:

$$p(i,i) = \frac{1}{2} \text{ and } p(i,i+1) = p(i,i-1) = \frac{1}{4},$$

i.e. half of the time the process does not move. In general, if P is the transition probability for a process, then $\widehat{P} = \frac{1}{2}(P + I)$ is the transition probability for the lazy version of the process. Recall the discussion on $\widehat{P} = \frac{1}{2}(P + I)$.

Let $n = 2^L$, where $L \in \mathbb{Z}_+$. So if T_1 is the first time the walker completes 1/4 of the circle, and T_2 is the amount of time after T_1 that it takes for the walker to complete 1/8 of the circle, and so on. Then

$$T = T_1 + T_2 + \cdots + T_{L-1} + 1$$

is a strong stationary time. Here

$$E[T] = E[T_1] + E[T_2] + \cdots + E[T_{L-1}] + 1 = \frac{1}{2} \cdot \left[\left(\frac{n}{4}\right)^2 + \left(\frac{n}{8}\right)^2 + \cdots + \left(\frac{n}{2^{L-1}}\right)^2 + 1\right]$$

and therefore

$$E[T] = \frac{n^2}{2} \cdot \sum_{j=2}^{L} \frac{1}{2^{2j}} \leq n^2 \cdot \sum_{j=2}^{\infty} \frac{1}{2^{2j+1}} .$$

2.3.2 Card shuffling examples and cutoff asymptotics.

Here we will explain some of the best known card shuffling examples.

Example. Random card to random location.

Example. [Aldous and Diaconis, 1981] Top-to-random card shuffling. (Also see Bremaud p225, and T.Lindvall) Here the strong stationary time is the time $\tau = \tau_n$ when the card that was the bottom card $\boxed{n}$ at time zero elevates to the top of the deck and then shuffled to the random location in the deck. One can show with the standard coupon collector argument that $E[\tau] = n \log n + O(n)$ and therefore $t_{mix} = O(n \log n)$. Indeed, let $T_{(1)}$ be the first time a card lands under $\boxed{n}$, $T_{(2)}$ be the second time a card lands under $\boxed{n}$, and so on. Then

$$\tau = T_{(1)} + T_{(2)} + \cdots + T_{(n-1)} + 1,$$

where each $T_{j)}$ is geometric with parameter $\frac{j}{n}$. Thus

$$E[\tau] = n \cdot \sum_{j=1}^{n} \frac{1}{j} = n \log n + O(n) \ .$$

However $t_n = n \log n$ is also a **cutoff time**: if ν_t denotes the distribution of X_t, the t-times shuffled deck, in S_n. Then

$$\|\nu_{(1-\epsilon)t_n} - \pi\|_{TV} \to 1 \quad \text{and } \|\nu_{(1+\epsilon)t_n} - \pi\|_{TV} \to 0 \quad \text{as } n \to \infty.$$

Indeed we have a coupon collector bound on τ: if V_j is the time the collection gets the jth coupon, $V_j \sim T_{(n-j+1)}$

$$\begin{aligned}
\|\nu_t - \pi\|_{TV} &\leq \mathbf{Prob}[\tau > t] \\
&= \mathbf{Prob}\left[\cup_{j=1}^{n}\{\text{coupon } \#j \text{ is not collected in } t \text{ drawings}\}\right] \\
&\leq \sum_{j=1}^{n} \mathbf{Prob}\left[\{\text{coupon } \#j \text{ is not collected in } t \text{ drawings}\}\right] \\
&= n\left(1 - \frac{1}{n}\right)^t \to 0
\end{aligned}$$

where $n\left(1 - \frac{1}{n}\right)^t \approx n^{-\epsilon} \to 0$ if $t = (1-\epsilon)t_n = (1-\epsilon)n \log n$. Thus showing

$$\|\nu_{(1+\epsilon)t_n} - \pi\|_{TV} \to 0 \quad \text{as } n \to \infty.$$

Moreover letting $t = n \log n + cn$ obtain

$$\|\nu_t - \pi\|_{TV} \leq \mathbf{Prob}[\tau > t] \leq n\left(1 - \frac{1}{n}\right)^t \approx e^{-c} = \frac{1}{4}$$

if $c = \log 4$. Thus $t_{mix} = n \log n + O(n)$.

Also $\|\nu_{(1-\epsilon)t_n} - \pi\|_{TV} \to 1$ as if only cards $\boxed{1}, \boxed{2}, \ldots, \boxed{k}$ took part in the t shuffles, then the probability measure ν_t is distributed among the $\frac{n!}{(n-k)!}$ combinations in S_n where card $\boxed{\text{n}}$ is under $\boxed{\text{n-1}}$ which in turn is under $\boxed{\text{n-2}}$, and so on up to card $\boxed{\text{k+1}}$.

Example. *Random transpositions.* [Diaconis] and [Diaconis and Shahshahani, 1981]

2.3.3 Mixing times via coupling.

Example. *Lazy random walk on a two-dimensional torus $S = \mathbb{Z}^2/n\mathbb{Z}^2$.* The original "lazy" process has transition probabilities

$$\begin{aligned} p([i,j],[i+1]) &= \frac{1}{8} \\ p([i,j],[i-1,j]) &= \frac{1}{8} \\ p([i,j],[i,j+1]) &= \frac{1}{8} \\ p([i,j],[i,j-1]) &= \frac{1}{8} \\ p([i,j],[i,j]) &= \frac{1}{2} \end{aligned}$$

We construct a coupled process to estimate the mixing time. Here are the transition probabilities for the coupled process in $S \times S$: Suppose $X_t = (i_1, j_1)$ and $Y_t = (i_2, j_2)$, where the coordinates are given mod n. Then X_{t+1} is determined according to

$$
\begin{aligned}
p([i_1,j_1,i_2,j_2],[i_1,j_1,i_2,j_2+1]) &= \frac{1}{8} \\
p([i_1,j_1,i_2,j_2],[i_1,j_1,i_2,j_2-1]) &= \frac{1}{8} \\
p([i_1,j_1,i_2,j_2],[i_1,j_1+1,i_2,j_2]) &= \frac{1}{8} \\
p([i_1,j_1,i_2,j_2],[i_1,j_1-1,i_2,j_2]) &= \frac{1}{8} \\
p([i_1,j_1,i_2,j_2],[i_1+1,j_1,i_2,j_2]) &= \frac{1}{8} \quad \text{if } i_1 \neq i_2 \\
p([i_1,j_1,i_2,j_2],[i_1-1,j_1,i_2,j_2]) &= \frac{1}{8} \quad \text{if } i_1 \neq i_2 \\
p([i_1,j_1,i_2,j_2],[i_1,j_1,i_2+1,j_2]) &= \frac{1}{8} \quad \text{if } i_1 \neq i_2 \\
p([i_1,j_1,i_2,j_2],[i_1,j_1,i_2-1,j_2]) &= \frac{1}{8} \quad \text{if } i_1 \neq i_2 \\
p([i_1,j_1,i_2,j_2],[i_1+1,j_1,i_2+1,j_2]) &= \frac{1}{8} \quad \text{if } i_1 = i_2 \\
p([i_1,j_1,i_2,j_2],[i_1-1,j_1,i_2-1,j_2]) &= \frac{1}{8} \quad \text{if } i_1 = i_2 \\
p([i_1,j_1,i_2,j_2],[i_1,j_1,i_2,j_2]) &= \frac{1}{4} \quad \text{if } i_1 = i_2 \ .
\end{aligned}
$$

In words, each time, with probability 1/2, one of the walkers moves and the other skips the move until T_1, the first time they are on the same vertical line. After T_1, they move simultaneously along the horizontal axis, and alternate the moves (according to coin toss) in the vertical direction. Thus $E[T_1] \le 2 \cdot \left(\frac{n}{2}\right)^2$ and the mean coupling time

$$E[T_{coupling}] \le n^2 \ .$$

Suppose $X_t \sim \nu_0 P^t$ and $Y_t \sim \mu_0 P^t$. If T is the coupling time for (X_t, Y_t), then

$$
\begin{aligned}
\|\nu_0 P^t - \mu_0 P^t\|_{TV} &= \frac{1}{2}\sum_{x \in S} |\mathbf{Prob}[X_t = x] - \mathbf{Prob}[Y_t = x]| \\
&= \frac{1}{2}\sum_{x \in S} |\mathbf{Prob}[X_t = x,\ T > t] - \mathbf{Prob}[Y_t = x,\ T > t]| \\
&\le \mathbf{Prob}[T > t].
\end{aligned}
$$

If we let $\mu_0 = \pi$ then this bounds $\|\nu_0 P^t - \pi P^t\|_{TV}$. In other words, the coupling time T is a **strong stationary time**.

If we have the general bound on $\mathbf{Prob}[T > t]$ for all μ_0, then we will have an upper bound for the mixing time:

$$\max_x \|\nu_0 P^t - x P^t\|_{TV} \ge \|\nu_0 P^t - \pi\|_{TV}$$

as $\sum_x \pi(x)\|\nu_0 P^t - xP^t\|_{TV} \geq \|\nu_0 P^t - \pi\|_{TV}$ by convexity of $f(x) = \|\nu_0 P^t - xP^t\|_{TV}$.

Example. *Random transpositions via simplest coupling.* Order $O(n^2)$.

Example. *Random adjacent transpositions via coupling.* [Aldous and Fill, Ch 4-3]; originally in [D.B.Wilson 1997] Sharp order $O(n^3 \log n)$.

2.3.4 Mixing and relaxation times.

More general definition of mixing time: fix $\epsilon \in [0, 1)$, then if $X_t \sim \nu_t = \nu_0 P_t$,

$$t_{mix}(\epsilon) := \inf \{t \ : \ \|\nu_t - \pi\|_{TV} \leq \epsilon, \quad \text{all } \nu_0\} \ .$$

Now, we know from Perron-Frobenious theorem that

$$P^t = \Pi + O(t^{m_2 - 1}\lambda_2^t),$$

where by our assumption, $\lambda_2 = \lambda_{(2)} > 0$. Recall that m_2 was the multiplicity of λ_2. We have shown that for $t = K \cdot \tau_{rlx}, \quad \lambda_2^t \leq e^{-K}$ if $m_2 = 1$. So,

$$\|\nu_0 P^t - \pi\|_{TV} = \|\nu_0 P^t - \nu_0 \Pi\|_{TV} \leq C \cdot e^{-K} = \epsilon$$

for all ν_0 if $K = \log \frac{C}{\epsilon}$. Thus

$$t_{mix}(\epsilon) \leq \log \frac{C}{\epsilon} \cdot \tau_{rlx}.$$

In fact, $C = \frac{1}{\min_{x \in S}\{\pi(x)\}}$ will work (see [Peres]). Conversely,

$$t_{mix}(\epsilon) \geq (\tau_{rlx} - 1) \log\left(\frac{1}{2\epsilon}\right) \ .$$

In order to show this we observe that any v_j $(j = 2, \ldots, r)$,

$$|\lambda_j^t v_j(x)| = |(P^t v_j^T)(x)| = \left|\sum_{y \in S} p_t(x, y)v_j(y) - \pi(y)v_j(y)\right| \leq 2\|v_j\|_\infty \cdot \|xP^t - \pi\|_{TV}$$

as $< \mathbf{1}, v_j >_\pi = 0$ for all $j = 2, \ldots, r$. Taking $x \in S$ that maximizes v_j, i.e. $v_j(x) = \|v_j\|_\infty$, we obtain

$$|\lambda_j|^t \leq 2\|xP^t - \pi\|_{TV} \leq 2 \max_{\nu_0 \in S} \|\nu_0 P^t - \pi\|_{TV} \leq 2\epsilon$$

if $t = t_{mix}(\epsilon)$. Hence $t_{mix}(\epsilon) \geq \frac{\log\left(\frac{1}{2\epsilon}\right)}{\log\left(\frac{1}{|\lambda_j|}\right)} \geq \frac{\log\left(\frac{1}{2\epsilon}\right)}{\left(\frac{1}{|\lambda_j|} - 1\right)} \geq (\tau_{rlx} - 1)\log\left(\frac{1}{2\epsilon}\right)$ as $|\lambda_{(2)}| \geq |\lambda_j|$.

2.4 Long run behavior of continuous time processes.

Aldous and Fill, Bremaud Ch.8 Taking independent exponential waiting times transforms a discrete time model into a continuous time one, where if the exponential parameter is one, the probabilities turn into rates.

Superposition. If X_λ is an exponential random variable with parameter $\lambda > 0$, and X_μ is an exponential random variable with parameter $\mu > 0$, then $Y = X_\lambda \wedge X_\mu$ is an exponential random variable with parameter $\lambda + \mu$. Check!

So if we are given two continuous time processes on the same discrete state space S. Then the superposition of the two processes is a continuous time process whose rates are the sums of the corresponding rates for the two processes, e.g. consider two Poisson processes N_t and M_t with rates $\lambda > 0$ and $\mu > 0$, then $N_t + M_t$ is the Poisson process with rate $\lambda + \mu$.

Thinning. Consider a geometric random variable with probability of success p and waiting times in between the trials distributed as independent exponential random variables with parameter $\lambda > 0$. Then it is an exponential random variable with parameter λp.

2.4.1 Example: Ising model and Glauber dynamics.

[Y.Peres]

$$\pi(\sigma) = \frac{1}{Z(\beta)} e^{\beta \cdot \sum_{u \sim v} \sigma(u)\sigma(v)} ,$$

where β is the reciprocal of the temperature, $\beta = \frac{1}{T}$, and $Z(\beta)$ is the normalization constant.

Glauber dynamics:

$$\mathbf{Prob}[\sigma_{new}(v) = 1] = \frac{e^{\beta \cdot \sum_{u:u \sim v} \sigma(u)}}{e^{-\beta \cdot \sum_{u:u \sim v} \sigma(u)} + e^{\beta \cdot \sum_{u:u \sim v} \sigma(u)}}.$$

Coupling.[Y.Peres] If $d(\cdot,\cdot)$ is a metric on S such that $d(i,j) \geq 1$ for $i \neq j$. If one constructs a coupling so that

$$E[d(X_{t+1}, Y_{t+1})] \leq e^{-\gamma} E[d(X_t, Y_t)] \quad \text{for some} \quad \gamma > 0,$$

then for $\nu_0(x) = \delta_i(x)$ and $\mu_0(x) = \delta_j(x)$,

$$\|\nu_0 P^t - \mu_0 P^t\|_{TV} \leq e^{-\gamma t} d(i,j) \leq e^{-\gamma t} \text{Diam}(S)$$

and

$$t_{mix}(\epsilon) \leq \gamma^{-1} \log \frac{\text{Diam}(S)}{\epsilon} .$$

2.5 Homework #2.

Problem 1. *Coupling in* $\mathbb{Z}^3$. Consider a lazy random walk on $S = \mathbb{Z}^3$, i.e. do nothing with probability $1/2$, or move to one of the six neighbor vertices with probability $\frac{1}{12}$ each.

Construct a successful coupling, i.e. a coupled process (X_t, Y_t) on $S \times S$ with finite coupling time: $\mathbf{Prob}[T_{coupling} < \infty] = 1$.

Problem 2. *Mixing time for lazy simple random walk on d-dimensional torus.* Consider a lazy random walk on $S = \mathbb{Z}^d / n\mathbb{Z}^d = (\mathbb{Z}^d \bmod n)$, a d-dimensional torus. Here the walker does nothing with probability $1/2$, or moves to one of the $2d$ neighbor vertices with probability $\frac{1}{4d}$ each. Show that the mixing time

$$t_{mix} = O(n^2)\ .$$

Problem 3. *Mixing time for the free dynamics on a graph with n vertices.* Consider a spin system on $\mathbb{Z}_n = \{1, 2, \ldots, n\}$ where each site is occupied by either $+1$ or -1. There is an exponential clock with parameter $p > 0$, when it rings, one of the n sites (chosen uniformly) changes the spin. (There is an equivalent definition: each site has an independent exponential clock with fixed rate $\frac{p}{n}$. When the corresponding clock rings, the site changes the sign of the spin.) Show that $t_{mix} = O(n \log n)$.

Chapter 3

Gibbs Fields

[Bremaud Ch.7] and [Grimmett]

Given a site space S, a collection of random variables indexed by S, $\{X_v\}_{v\in S}$ is called a *random field*.

3.1 Markov random fields.

Let S be a graph, then each vertex $v \in S$ has a neighborhood system $N_v = \{u \in S \ : \ u \sim v\}$, where $u \sim v$ denotes two vertices connected by an edge.

Definition. The random field $\{X_v\}_{v\in S}$ is called a **Markov random field** if for any vertex $v \in S$, X_v is independent of $\{X_w : \ w \in S \setminus (v \cup N_v)\}$ when conditioned on the values $\{X_u : \ u \in N_v\}$.

Example. *Ising model on 2-D torus $\mathbb{Z}^2/n\mathbb{Z}^2$.* Here the Gibbs potential is given by the following Hamiltonian:

$$\mathcal{H}(\sigma) = -\frac{1}{2} \sum_{u,v:\ u\sim v} \sigma(u)\sigma(v) = \quad - \sum_{edges\ \ e=[u,v]} \sigma(u)\sigma(v)$$

and probability of a configuration $\sigma \in \Lambda^S = \{-1,+1\}^S$ is

$$\pi(\sigma) = \frac{e^{-\beta\mathcal{H}(\sigma)}}{Z(\beta)}, \quad \text{where} \quad \beta = \frac{1}{T}$$

and $Z(\beta) = \sum_{\sigma\in\Lambda} e^{-\beta\mathcal{H}(\sigma)}$ is the normalizing factor. The Hamiltonian $\mathcal{H}(\sigma)$ can be expressed through local Hamiltonians: for each $v \in S$, we define the local Hamiltonian

$$\mathcal{H}_{local}(\sigma, v) = - \sum_{u:\ u\sim v} \sigma(u)\sigma(v) \ .$$

Then

$$\mathcal{H}(\sigma) = \frac{1}{2}\sum_{v\in S} \mathcal{H}_{local}(\sigma, v) \ .$$

Glauber dynamics: Each time we randomly pick a vertex $v \in G$, and erase the spin $\sigma(v)$ at v. Let σ_+ (respectively σ_-) be the configuration we get if we place $\sigma(v) = +1$ (respectively $\sigma(v) = -1$) spin at v. Then the probability of switching from σ to σ_+ is given by

$$\mathbf{Prob}(\sigma \to \sigma_+) = \frac{e^{-\beta\mathcal{H}(\sigma_+)}}{e^{-\beta\mathcal{H}(\sigma_-)} + e^{-\beta\mathcal{H}(\sigma_+)}} = \frac{e^{-\beta\mathcal{H}_{local}(\sigma_+,v)}}{e^{-\beta\mathcal{H}_{local}(\sigma_-,v)} + e^{-\beta\mathcal{H}_{local}(\sigma_+,v)}} \ .$$

So Glauber dynamics is a random walk on the space of all configurations Λ^S such that the probability measure π is stationary w.r.t. Glauber dynamics. Moreover, if $\tanh(\beta) < \frac{1}{4}$, the mixing time $t_{mix} = O(n \log n)$. Thus the Glauber dynamics is a fast way to generate π. Glauber dynamics is an important example of a **Gibbs sampler**.

The **Ising model with external field** has Hamiltonian defined as

$$\mathcal{H}(\sigma) = -\frac{1}{2}\sum_{u,v:\ u\sim v} \sigma(u)\sigma(v) - h\cdot\sum_v \sigma(v) = \quad -\sum_{edges\ \ e=[u,v]} \sigma(u)\sigma(v) - h\cdot\sum_v \sigma(v)$$

Example. *Ising model on 1-D torus $S = \mathbb{Z}/n\mathbb{Z}$.* [Bremaud, Ch 7, Example 1.4], [Baxter]

$$Z(\beta) = \sum_{\sigma\in\Lambda^S} e^{\beta\sum_{j=1}^n \sigma(j)\sigma(j+1)} = \sum_{\sigma\in\Lambda^S} R(\sigma(1),\sigma(2))R(\sigma(2),\sigma(3))\dots R(\sigma(n),\sigma(1)),$$

where $R(x,y) = e^{\beta xy}$ for all $x, y \in \Lambda = \{-1,+1\}$. The transformation matrix

$$R = \begin{pmatrix} R(-1,-1) & R(-1,+1) \\ R(+1,-1) & R(+1,+1) \end{pmatrix} = \begin{pmatrix} e^{\beta} & e^{-\beta} \\ e^{-\beta} & e^{\beta} \end{pmatrix}$$

generates $Z(\beta)$:

$$Z(\beta) = \sum_{\sigma(1)\in\Lambda} R^n(\sigma(1),\sigma(1)) = R^n(-1,-1) + R^n(+1,+1) = Tr(R^n) = \lambda_{R1}^n + \lambda_{R2}^n,$$

where $\lambda_{R1} = e^{\beta} + e^{-\beta}$ and $\lambda_{R2} = e^{\beta} - e^{-\beta}$ are the two eigenvalues of matrix R.

3.1.1 Gibbs-Markov equivalence.

Definition. A subset $C \subset S$ is called a *clique* for all pairs of sites $u, v \in C$ $(u \neq v)$, u and v must be neighbors.

Definition. The set of random variables $\sigma = \{X_v\}_{v\in S}$ is a *Gibbs field* if the joint distribution function can be written as

$$\pi(\sigma) = \frac{1}{Z_T} e^{-\frac{1}{T}\mathcal{E}(\sigma)} \qquad \text{for some } T > 0$$

with the energy function $\mathcal{E} : \Lambda^S \to \mathbb{R} \cup \{+\infty\}$ defined via the Gibbs potential functions $\{V_C(\sigma)\}_{C \subset S}$ in the following way:

$$\mathcal{E}(x) = \sum_{C \subset S} V_C(\sigma) ,$$

where $V_C : \Lambda^S \to \mathbb{R} \cup \{+\infty\}$ satisfy

- $V_C \equiv 0$ if C is not a clique;
- $V_C(\sigma)$ depends entirely on values of σ at C, $\sigma(C)$: if $\sigma_1, \sigma_2 \in \Lambda^S$ such that $\sigma_1(v) = \sigma_2(v)$ for all $v \in C$, then
$$V_C(\sigma_1) = V_C(\sigma_2) .$$

Theorem 11. *Gibbs fields are Markov fields.*

Assume a positivity condition. Then the following is true.

Theorem 12. Hammersley-Clifford Theorem. *Markov fields are Gibbs fields.*

Proof (Grimmett, 1973): Let 0 be one of the states in Λ. Denote by $\overrightarrow{0}$ the combination in Λ^S with zero states at all sites, i.e. $\overrightarrow{0}(v) = 0$ for any $v \in S$. Also, for any subset $B \subset S$, let $\sigma^B \in \Lambda^S$ denote the configuration such that $\sigma^B \equiv \sigma$ on B and $\sigma^B \equiv 0$ on $S \setminus B$. Let us define

$$V_A(\sigma) = \sum_{B \subset A} (-1)^{|A-B|} \log \frac{\pi(\overrightarrow{0})}{\pi(\sigma^B)} .$$

Then, by the Möbius formula (see Brémaude, p.262),

$$\log \frac{\pi(\overrightarrow{0})}{\pi(\sigma^A)} = \sum_{B \subset A} V_B(\sigma) .$$

Thus, taking $A = S$, obtain the expression for the MRF distribution π:

$$\pi(\sigma) = \pi(\overrightarrow{0}) e^{-\sum_{B \subset S} V_B(\sigma)} .$$

Now, since π is the distribution of a Markov random field, for any subset $B \subset S$ and site $v \notin B$,

$$\pi(\sigma^B) = \pi(\sigma^B(v) |\, \sigma^B(S \setminus v)) \cdot \pi(\sigma^B(S \setminus v)) = \pi(\sigma^B(v) |\, \sigma^B(N_v)) \cdot \pi(\sigma^B(S \setminus v))$$

and

$$\pi(\sigma^{B+v}) = \pi(\sigma^{B+v}(v) |\, \sigma^{B+v}(S \setminus v)) \cdot \pi(\sigma^{B+v}(S \setminus v)) = \pi(\sigma^{B+v}(v) |\, \sigma^{B+v}(N_v)) \cdot \pi(\sigma^{B+v}(S \setminus v)) ,$$

where $\pi(\sigma^{B+v}(S \setminus v)) = \pi(\sigma^B(S \setminus v))$ as $\sigma^{B+v}(S \setminus v) = \sigma^B(S \setminus v)$. Thus for any $B \subset S$ and $v \notin B$,

$$\frac{\pi(\sigma^{B+v})}{\pi(\sigma^B)} = \frac{\pi(\sigma^{B+v}(v)|\ \sigma^{B+v}(N_v))}{\pi(\sigma^B(v)|\ \sigma^B(N_v))} \tag{3.1}$$

Now, we need to show that $V_A \equiv 0$ if A is not a clique. Suppose A is not a clique, then there is a pair of sites $u, v \in A$ that are not neighbors. Then

$$\begin{aligned}
V_A(\sigma) &= \sum_{B \subset A} (-1)^{|A-B|} \log \frac{\pi(\overrightarrow{0})}{\pi(\sigma^B)} \\
&= \sum_{B \subset A-u-v} (-1)^{|A-B|} \log \frac{\pi(\overrightarrow{0})}{\pi(\sigma^B)} + \sum_{B \subset A-u-v} (-1)^{|A-B-u|} \log \frac{\pi(\overrightarrow{0})}{\pi(\sigma^{B+u})} \\
&+ \sum_{B \subset A-u-v} (-1)^{|A-B-v|} \log \frac{\pi(\overrightarrow{0})}{\pi(\sigma^{B+v})} + \sum_{B \subset A-u-v} (-1)^{|A-B-u-v|} \log \frac{\pi(\overrightarrow{0})}{\pi(\sigma^{B+u+v})} \\
&= \sum_{B \subset A-u-v} (-1)^{|A-B|} \log \frac{\pi(\sigma^{B+u})\pi(\sigma^{B+v})}{\pi(\sigma^B)\pi(\sigma^{B+u+v})} \\
&= \sum_{B \subset A-u-v} (-1)^{|A-B|} \log \frac{\pi(\sigma^{B+u}|\sigma^{B+u}(N_v))\ \pi(\sigma^{B+v}|\sigma^{B+v}(N_v))}{\pi(\sigma^B|\sigma^B(N_v))\ \pi(\sigma^{B+u+v}|\sigma^{B+u+v}(N_v))} \quad \text{by eq. (3.1) above} \\
&= 0
\end{aligned}$$

for all $\sigma \in \Lambda^S$ as $\pi(\sigma^{B+u}|\sigma^{B+u}(N_v)) = \pi(\sigma^B|\sigma^B(N_v))$ and $\pi(\sigma^{B+u+v}|\sigma^{B+u+v}(N_v)) = \pi(\sigma^{B+v}|\sigma^{B+v}(N_v))$ since $u \notin \{v\} \cup N_v$.

Also, V_A was defined so that $V_A(\sigma)$ depends entirely on values of σ at A, $\sigma(A)$

□

Example. *Markov chain over* $[0, n]$ *time interval. Gibbs sampler.*

3.2 Ising model and bond percolation.

3.2.1 Spin systems

3.2.2 Phase transition.

3.2.3 Criticality.

3.2.4 Two dimensional bond percolation: $p_c = \frac{1}{2}$.

3.3 Monte Carlo Markov Chain method, Metropolis Algorithm and Gibbs sampler.

Bremaud + [N.Madras, "Lecture on Monte Carlo Methods"]

Bibliography

[1] D.Aldous and J.A.Fill, *Reversible Markov Chains and Random Walks on Graphs.* http://www.stat.berkeley.edu/users/aldous/

[2] P.Brémaud, *MARKOV CHAINS: GIBBS FIELDS, MONTE CARLO SIMULATION, AND QUEUES.* Springer-Verlag (1998)

[3] P.Diaconis and M.Shahshahani, *Generating a random permutation with random transpositions* Z. Wahrsch. Verw. Gebiete **57** (1981), 159-179.

[4] P.Diaconis, *Group Representations in Probability and Statistics.* Institute of Mathematical Statistics, Hayward CA (1973).

[5] Y.Peres, *Probability on Trees: An Introductory Climb.* http://www.stat.berkeley.edu/users/peres

[6] Y.Peres, *Mixing for Markov Chains and Spin Systems.* http://www.stat.berkeley.edu/users/peres/ubc.pdf