A Brief(?) Introduction to the Markov Chain Monte Carlo Method

Russell Martin

October 20, 2005
The MCMC Paradigm

$\Omega$ is a (typically finite) set.

- Graph colourings of $G$ (proper, Ising and Potts models)
- Contingency tables
- Perfect and near-perfect matchings in a bipartite graph (Permanent of a 0-1 matrix)
- Linear extensions of a partial order
The MCMC Paradigm

$\Omega$ is a (typically finite) set.
- Graph colourings of $G$ (proper, Ising and Potts models)
- Contingency tables
- Perfect and near-perfect matchings in a bipartite graph
  (Permanent of a 0-1 matrix)
- Linear extensions of a partial order

We’re interested in sampling from $\Omega$ according to some distribution and/or finding $|\Omega|$.
Ω is a (typically finite) set.

- Graph colourings of G (proper, Ising and Potts models)
- Contingency tables
- Perfect and near-perfect matchings in a bipartite graph (Permanent of a 0-1 matrix)
- Linear extensions of a partial order

We’re interested in sampling from Ω according to some distribution and/or finding |Ω|.

In most “interesting” cases computing |Ω| exactly is hard (i.e. \#P-complete), so we settle for a good approximation.
The MCMC Paradigm

\( \Omega \) is a (typically finite) set.
- Graph colourings of G (proper, Ising and Potts models)
- Contingency tables
- Perfect and near-perfect matchings in a bipartite graph (Permanent of a 0-1 matrix)
- Linear extensions of a partial order

We’re interested in sampling from \( \Omega \) according to some distribution and/or finding \(|\Omega|\).

In most “interesting” cases computing \(|\Omega|\) exactly is hard (i.e. \( \#P \)-complete), so we settle for a good approximation.

How do we do this sampling and (approximate) counting?
The MCMC Paradigm (cont.)

$H$ is a directed graph with $V(H) = \Omega$.
(e.g. the vertices of $H$ are the proper colourings of $G$)

Edges in $E(H)$ represent “local moves” that connect pairs in $\Omega$.
(e.g. two proper graph colourings of $G$ that differ at a single vertex)

Design a Markov chain $\mathcal{M}$ that is a random walk on $H$. 

$P$ is the transition matrix of this chain.

1. $p_{ij} = \text{Prob}(i \rightarrow j \text{ in one step of } \mathcal{M})$

2. $p_{ij} \geq 0 \quad \forall i, j \in \Omega$

3. $\sum_j p_{ij} = 1 \quad \forall i \in \Omega$ ($P$ is row stochastic.)

4. $(e, f) \in E(G) \iff p_{ef} > 0$

5. $p(n)_{ij} = \text{Prob}(i \rightarrow j \text{ in exactly } n \text{ steps of } \mathcal{M})$
The MCMC Paradigm (cont.)

$H$ is a directed graph with $V(H) = \Omega$. (e.g. the vertices of $H$ are the proper colourings of $G$)

Edges in $E(H)$ represent “local moves” that connect pairs in $\Omega$. (e.g. two proper graph colourings of $G$ that differ at a single vertex)

Design a Markov chain $\mathcal{M}$ that is a random walk on $H$.

$P$ is the transition matrix of this chain.

1. $p_{ij} = \text{Prob}(i \rightarrow j \text{ in one step of } \mathcal{M})$
2. $p_{ij} \geq 0 \ \forall i, j \in \Omega$
3. $\sum_j p_{ij} = 1 \ \forall i \in \Omega$ ($P$ is row stochastic.)
4. $e = (i, j) \in E(G) \iff p_{ij} > 0$
5. $p_{ij}^{(n)} = \text{Prob}(i \rightarrow j \text{ in exactly } n \text{ steps of } \mathcal{M})$
Assuming some mild conditions on $\mathcal{M}$, the Markov chain converges to a unique limiting distribution. These conditions are:

- **Irreducibility** ($H$ is strongly connected.)
- **Aperiodicity** ($H$ isn’t bipartite, etc.)

The limiting, or stationary, distribution $\pi$ is the unique normalized vector that satisfies:

- $\pi_i \geq 0 \ \forall i \in \Omega$,
- $\pi P = \pi$,
- and $\lim_{n \to \infty} p_{ij}^{(n)} = \pi_j \ \forall i, j \in \Omega$. 
The Big Question...

For how many steps do we have to iterate the chain until we’re close to $\pi$?

**total variation distance**

$$d_{tv}(P^{(n)}, \pi) = \frac{1}{2} \max_{i \in \Omega} \sum_j |p_{ij}^{(n)} - \pi_j|$$

**mixing time** For $\varepsilon > 0$, we define

$$\tau(\varepsilon) = \min\{t | d_{tv}(P^{(n)}, \pi) \leq \varepsilon \text{ for all } n \geq t\}. $$
We hope to prove that the chain is *rapidly mixing*, which means that

$$\tau(\varepsilon) \leq \text{poly}(n, \log(1/\varepsilon)),$$

where $n$ is the size of the input (e.g. the number of vertices in the input graph $G$ that we wish to colour).
We usually work with *reversible* Markov chains, i.e. those that satisfy

\[\pi_i \, p_{ij} = \pi_j \, p_{ji} \quad \text{for all } i, j \in \Omega.\]

This implies the eigenvalues of $P$ are real numbers, and for an ergodic chain, the Perron-Frobenius theorem tells us that

\[1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_{|\Omega|} > -1.\]
We usually work with *reversible* Markov chains, i.e. those that satisfy
\[ \pi_i p_{ij} = \pi_j p_{ji} \quad \text{for all } i, j \in \Omega. \]

This implies the eigenvalues of \( P \) are real numbers, and for an ergodic chain, the Perron-Frobenius theorem tells us that
\[ 1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_{|\Omega|} > -1. \]

**Theorem (Sinclair (1993), et. al.)**

Let \( \text{Gap}(P) = 1 - \max\{\lambda_2, |\lambda_{|\Omega|}|\} \) and \( \pi_{\min} = \min_j \pi_j \).

1. \[ \tau(\varepsilon) \leq \frac{1}{\text{Gap}(P)} \log\left(\frac{1}{\varepsilon \pi_{\min}}\right) \]
2. \[ \tau(\varepsilon) \geq \frac{|\lambda_2|}{2\text{Gap}(P)} \log\left(\frac{1}{2\varepsilon}\right) \]
The Big Problem...

\[ P \text{ is huge!!} \]

Calculating its eigenvalues seems (and often is) really hard!

(Recall that often we don’t even know the exact size of \( P \).)
MCMC methods that (sometimes) help

Coupling and path coupling
Conductance (isoperimetric inequalities) and “congestion”
Comparison of Markov chains
Markov chain decomposition
Evolving sets
Log-Sobolev inequalities

Coupling from the Past
  (Exact samples, no a priori bounds on time required).
Let $G$ be some fixed graph, and let $q$ denote a positive integer. 
$\#G =$ number of proper $q$-colourings of $G$.

Suppose we have a sampler that can return proper $q$-colourings of a graph (almost) uniformly at random.

How can we use this to (approximately) find $\#G$?
The Connection between Sampling and Counting

\[ |E(G)| = m \]

Define a sequence of graphs

\[ G = G_m \supset G_{m-1} \supset \cdots \supset G_1 \supset G_0 = \text{empty graph (no edges)} \]

obtained by deleting the edges of \( G \) in some order.
The Connection between Sampling and Counting

\[ |E(G)| = m \]

Define a sequence of graphs

\[ G = G_m \supset G_{m-1} \supset \cdots \supset G_1 \supset G_0 = \text{empty graph (no edges)} \]

obtained by deleting the edges of \( G \) in some order.

Then,

\[ \#G = \frac{\#G_m}{\#G_{m-1}} \cdot \frac{\#G_{m-1}}{\#G_{m-2}} \cdots \frac{\#G_1}{\#G_0} \cdot \#G_0, \]

where \( \#G_0 = q^n \).

Use the sampler to estimate \( \frac{\#G_i}{\#G_{i-1}} \) for each \( i \) (use Chernoff bounds, etc).
A Markovian coupling for $\mathcal{M}$ is a stochastic process $(X_t, Y_t)$ on $\Omega \times \Omega$ such that

1. each process $(X_t)$ and $(Y_t)$, viewed in isolation, is faithful to $\mathcal{M}$;

2. if $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$.
A Markovian coupling for $\mathcal{M}$ is a stochastic process $(X_t, Y_t)$ on $\Omega \times \Omega$ such that

1. each process $(X_t)$ and $(Y_t)$, viewed in isolation, is faithful to $\mathcal{M}$;
2. if $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$.

**Lemma (Aldous)**

$$d_{tv}(P^{(t)}, \pi) \leq \max_{x,y \in \Omega} Pr(X_t \neq Y_t \mid X_0 = x, Y_0 = y)$$
Path Coupling

Bubley and Dyer [1997]

Define a coupling on a subset $S$ of $\Omega \times \Omega$.

$d =$ an integer-valued distance function on $S$.

Extend the coupling, via shortest paths, to a coupling on $\Omega \times \Omega$. 
Path Coupling

Bubley and Dyer [1997]

Define a coupling on a subset $S$ of $\Omega \times \Omega$. $d$ = an integer-valued distance function on $S$.

Extend the coupling, via shortest paths, to a coupling on $\Omega \times \Omega$. Show (we hope!) the coupling is contracting on $S$, i.e.

$$\mathbb{E}(\Delta d(X, Y)) \leq 0 \text{ for all pairs } (X, Y) \in S.$$

Show there’s some variance, i.e.

$$\text{Pr}(d(X_{t+1}, Y_{t+1}) \neq d(X_t, Y_t)|X_t \neq Y_t) \text{ is “big” for all } X_t \neq Y_t.$$ 

Conclude a mixing bound on the chain.
Sampling Proper Graph Colourings

$q$ colours.
$G$, a fixed graph on $n$ vertices, max degree $\delta$.

Markov chain is single-site “heat-bath” dynamics. That is, each step of the chain consists of the following steps:

- Pick a vertex $v \in G$ uniformly at random.
- Pick a colour $c \in \{1 \ldots q\}$ uniformly at random.
- Try to recolour $v$ using $c$ if this gives a proper colouring, otherwise do nothing.

$S =$ set of pairs of colourings of $G$ that differ at a single vertex.

$d(X, Y) =$ Hamming distance.

(Must extend $\Omega$ to include all, proper and improper, colourings of $G$, but the improper colourings are “transient” states.)
Jerrum (1995)

\[ d(X, Y) = 1 \]
Introduction
Graph Colourings
A Curious(?) State of Affairs

Sampling Proper Graph Colourings

[Graph showing two graphs with one marked as red]
Sampling Proper Graph Colourings

A Curious(?) State of Affairs

The MCMC Method
Sampling Proper Graph Colourings

The MCMC Method
There’s at least $q - \delta$ ways to get the colourings to agree.
Sampling Proper Graph Colourings

R. Martin  The MCMC Method
Sampling Proper Graph Colourings
Sampling Proper Graph Colourings

(red, green)
Sampling Proper Graph Colourings

(green, red)
There’s at most $\delta$ ways to increase the distance (by one).
Putting this all together, we have

\[ E(\Delta d(X, Y)) \leq -\frac{(q - \delta)}{qn} + \frac{\delta}{qn} \leq 0 \text{ if } q \geq 2\delta. \]

Rapid mixing in the case that \( q \geq 2\delta \).

\[ \tau(\varepsilon) \in O\left(n \log \left(\frac{n}{\varepsilon}\right)\right) \text{ if } q > 2\delta. \]

\[ \tau(\varepsilon) \in O\left(n^3 \log \left(\frac{1}{\varepsilon}\right)\right) \text{ if } q = 2\delta. \]
The Graph Colouring “Industry”

Jerrum [1995] Rapid Mixing (R.M.) of Glauber dynamics when $q \geq 2\delta$

Salas, Sokal [1997] $q = 7$ on $\mathbb{Z}^2$ (C.A.)
$q = 4, 5$ on hexagonal lattice (C.A.)
$q = 11$ on triangular lattice (C.A.)

Vigoda [2000] R.M. when $q \geq \frac{11}{6}\delta$


Luby, Randall, Sinclair [2001] 3-col. of $\mathbb{Z}^2$ with fixed boundary
The Graph Colouring “Industry”

Kenyon, Mossel, Peres [2001] R.M. and exp. decay

Aclioptas, Molloy, Moore, van Bussel [2004] $q = 6$ on $\mathbb{Z}^2$ (C.A.)

Hayes, Vigoda [2003, 2004] $q > \alpha \cdot \delta, \delta = \Omega(\log n), \text{girth} \geq 4$
$q > (1 + \varepsilon)\delta, \delta = \Omega(\log n), \text{girth} \geq 9$
$\alpha^\alpha = e$ that is, $\alpha \approx 1.76322$
The Graph Colouring “Industry”

Dyer, Frieze, Hayes, Vigoda [2004]
\[ q > \max(\alpha \delta, C), \text{girth } \geq 6 \]

Weitz [2004] Connections between Strong Spatial Mixing (S.S.M.) and R.M.

Goldberg, Martin, Paterson [2004]
3-col. of rectangular regions of \( \mathbb{Z}^2 \)
\textit{without} a fixed boundary
S.S.M. for triangle-free graphs with
\[ q > \alpha \delta - \gamma \text{ colours } (\gamma \approx 0.47) \]
R.M. above for “neighbourhood-amenable” graphs
R.M. for \( q = 10 \) on \( \mathbb{Z}^3 \) and triangular lattice
Consider colouring finite (say, rectangular) portions of the lattice $\mathbb{Z}^2$.

$q \geq 8$  R.M. by Jerrum’s result.
$q = 7$  Bubley, Dyer, Greenhill, Jerrum (C.A.)
Goldberg, Martin, Paterson (“hand proof”)
$q = 6$  Aclioptas, Molloy, Moore, van Bussel (C.A.)
Goldberg, Martin, Paterson (C.A.)

$q = 3$  Luby, Randall, Sinclair (fixed boundary)
Goldberg, Martin, Paterson (free boundary)

$q = 2$  Only two colourings!
Consider colouring finite (say, rectangular) portions of the lattice $\mathbb{Z}^2$.

$q \geq 8$ R.M. by Jerrum’s result.
$q = 7$ Bubley, Dyer, Greenhill, Jerrum (C.A.)
Goldberg, Martin, Paterson (“hand proof”)
$q = 6$ Aclioptas, Molloy, Moore, van Bussel (C.A.)
Goldberg, Martin, Paterson (C.A.)
$q = 5$ ???
$q = 4$ ???
$q = 3$ Luby, Randall, Sinclair (fixed boundary)
Goldberg, Martin, Paterson (free boundary)
$q = 2$ Only two colourings!