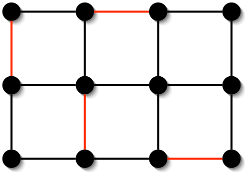
A tutorial on efficient sampling

Mark Jerrum
School of Informatics
University of Edinburgh

BCTCS, Swansea, 5th April 2006

Example 1: Matchings (monomer-dimer)

Instance: a graph $G = (V, E)$.



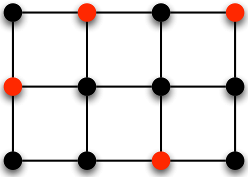
A *matching* is a collection $M \subseteq E$ of vertex-disjoint edges.

$$\pi(M) = \lambda^{|M|}/Z, \quad \text{where } Z = \sum_{M} \lambda^{|M|}.$$

Task: Sample from π, efficiently (certainly in time polynomial in $n = |V|$).

Example 2: Independent sets (hard-core gas)

Instance: a graph $G = (V, E)$.



An *independent set* is a subset $I \subseteq V$ of non-adjacent vertices.

$$\pi(I) = \lambda^{|I|}/Z', \quad \text{where } Z' = \sum_I \lambda^{|I|}.$$

Task: As before.

Computational complexity

- Despite their similarity, one of these two sampling problems is tractable and the other intractable.
- They are both trivial as decision problems.
- They are both hard (#P-complete) as counting problems.
- Approximate counting is strongly related to sampling. So one is tractable as an approximate counting problem and the other intractable.

Let's dive in fearlessly, using matching as an example.

Sequential choice

For convenience assume $\lambda = 1$.

- $M := \emptyset$.
- For each edge $e \in E(G)$ in turn $(*)$:
  - If e is “blocked” do nothing.
  - If e is “free”, add it to M with probability $\frac{1}{2}$.

The resulting distribution is highly dependent on the order $(*)$.

Sequential choice

For convenience assume $\lambda = 1$.

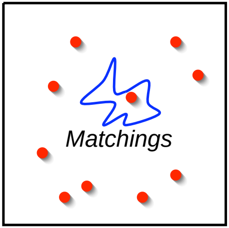
- $M := \emptyset$.
- For each edge $e \in E(G)$ in turn $(*)$:
  - If e is "blocked" do nothing.
  - If e is "free", add it to M with probability $\frac{1}{2}$.

The resulting distribution is highly dependent on the order $(*)$.

Example

For a path on n vertices, the asymptotic density of edges in the resulting matching is $\frac{1}{3}$, as against the correct $\frac{1}{2}\left(1 - 1/\sqrt{5}\right) = 0.276+$.

Monte Carlo (Dart throwing)

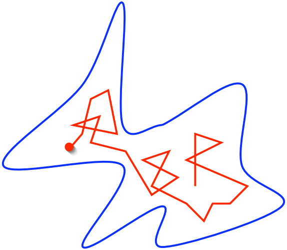


All subsets of E

- Until success:
    - Choose $M \subseteq E$ u.a.r.
    - If M is a matching, output M.

Correct distribution, but exponential running time.

Markov chain Monte Carlo



- Repeat:
    - Choose $e \in E$ u.a.r.
    - If e is blocked, do nothing.
    - Otherwise:
        - with probability $\frac{1}{2}$, $M := M \setminus \{e\}$, or
        - with probability $\frac{1}{2}$, $M := M \cup \{e\}$.

Mixing time

The trial just described defines the transition probabilities P of a Markov chain on state space

$$\Omega = \{\text{All matchings in } G\}.$$

The Markov chain is irreducible and aperiodic, and its stationary distribution π is uniform.

We are interested in the *mixing time* τ of the Markov chain, i.e., the time to convergence to near stationarity:

$$\tau = \max_{x \in \Omega} \min \left\{ t : \|P^t(x, \cdot) - \pi\|_{\mathrm{TV}} \leq e^{-1} \right\},$$

where $\|\sigma\|_{\mathrm{TV}} = \frac{1}{2} \sum_{x \in \Omega} |\sigma(x)|$.

Canonical paths/Multi-commodity flow

For every pair of states $x, y \in \Omega$, define a *canonical path* γ_{xy} from x to y using valid transitions of the MC.
"Congestion constant" ϱ:

$$\sum_{\gamma_{xy} \ni (z,z')} \pi(x)\pi(y)\,|\gamma_{xy}| \leq \varrho\,\pi(z)P(z,z'), \quad \forall z, z'.$$

Canonical paths/Multi-commodity flow

For every pair of states $x, y \in \Omega$, define a *canonical path* γ_{xy} from x to y using valid transitions of the MC.
"Congestion constant" ϱ:

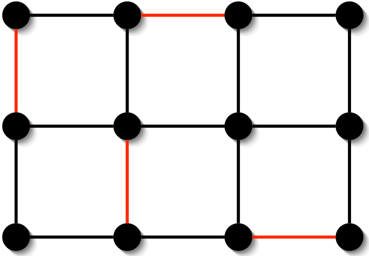
$$\sum_{\gamma_{xy} \ni (z,z')} \pi(x)\pi(y)\,|\gamma_{xy}| \leq \varrho\,\pi(z)P(z,z'), \quad \forall z, z'.$$

Theorem (Diaconis, Stroock; Sinclair)

$$\tau = O(\varrho \log \pi_{\min}^{-1}).$$

Richer set of transitions

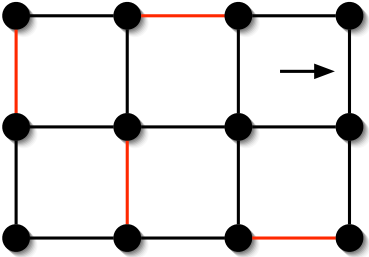
Convenient to augment existing “add” and “delete” transitions with a “displace”:



[Broder, 1986; J. & Sinclair, 1988]

Richer set of transitions

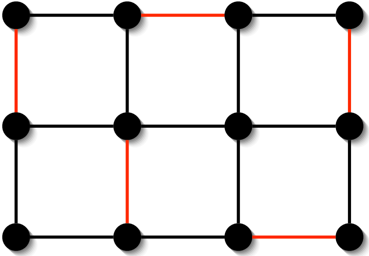
Convenient to augment existing “add” and “delete” transitions with a “displace”:



[Broder, 1986; J. & Sinclair, 1988]

Richer set of transitions

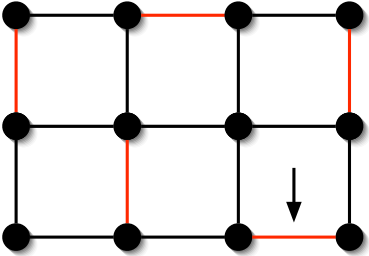
Convenient to augment existing "add" and "delete" transitions with a "displace":



[Broder, 1986; J. & Sinclair, 1988]

Richer set of transitions

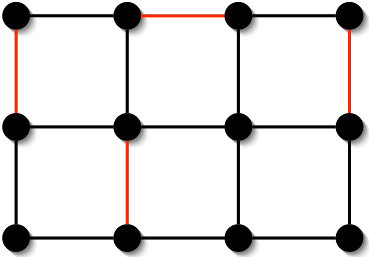
Convenient to augment existing “add” and “delete” transitions with a “displace”:



[Broder, 1986; J. & Sinclair, 1988]

Richer set of transitions

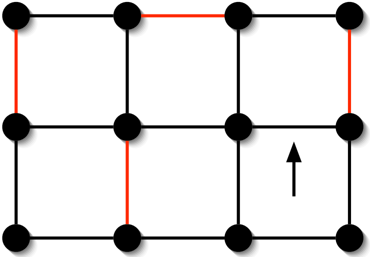
Convenient to augment existing "add" and "delete" transitions with a "displace":



[Broder, 1986; J. & Sinclair, 1988]

Richer set of transitions

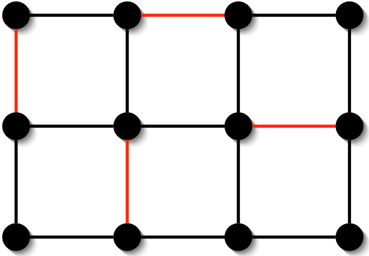
Convenient to augment existing “add” and “delete” transitions with a “displace”:



[Broder, 1986; J. & Sinclair, 1988]

Richer set of transitions

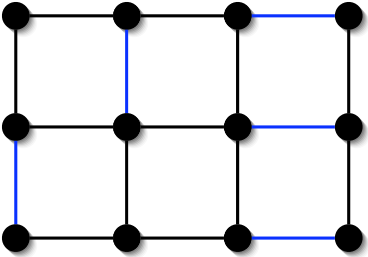
Convenient to augment existing “add” and “delete” transitions with a “displace”:



[Broder, 1986; J. & Sinclair, 1988]

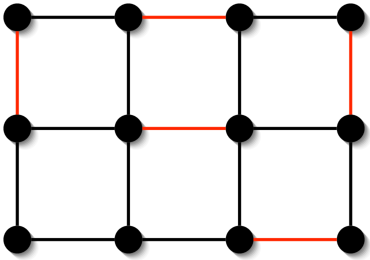
Canonical paths for matchings

To get from the blue matching. . .



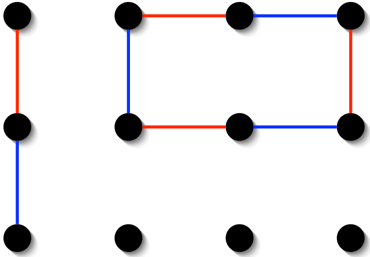
Canonical paths for matchings

. . . to the red matching. . .



Canonical paths for matchings

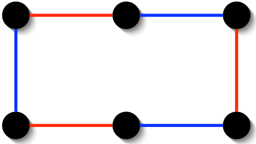
. . . first superimpose red and blue (symmetric difference). . .



and then "unwind" each component (path or cycle).

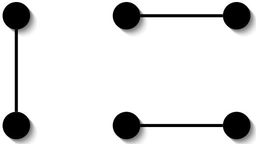
“Unwinding” a cycle

The cycle:



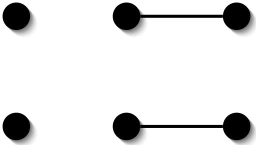
"Unwinding" a cycle

Initial matching:



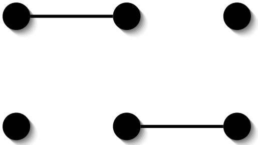
"Unwinding" a cycle

After 1 step:



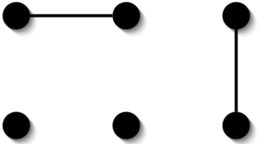
“Unwinding” a cycle

After 2 steps:



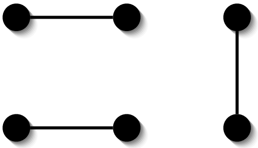
“Unwinding” a cycle

After 3 steps:



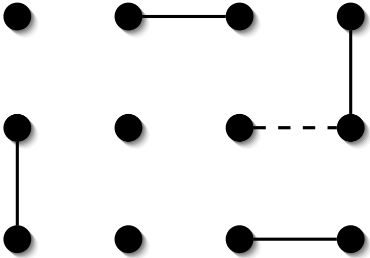
"Unwinding" a cycle

After 4 steps (final matching):



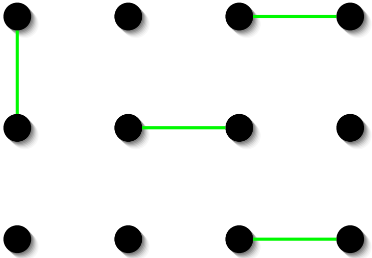
Encoding a canonical path through a transition

A transition:



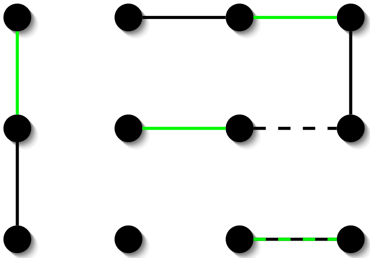
Encoding a canonical path through a transition

An encoding (matching):



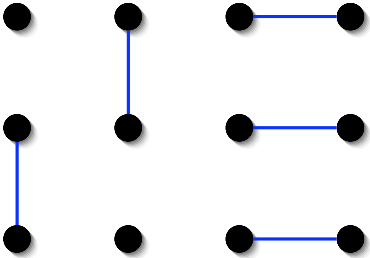
Encoding a canonical path through a transition

Superposition reveals the initial and final matching:



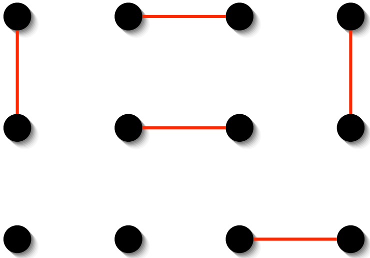
Encoding a canonical path through a transition

Superposition reveals the initial and final matching:



Encoding a canonical path through a transition

Superposition reveals the initial and final matching:



Calculating the congestion

The encoding argument shows that the number of canonical paths passing through a given transition is roughly equal to the size of the state space.

Pursuing the calculation in more detail yields:

Theorem (J. & Sinclair)

$\varrho = O(nm\bar{\lambda}^2)$, *where* $n = |V|$, $m = |E|$ *and* $\bar{\lambda} = \max\{\lambda, 1\}$.

Corollary

$\tau = O(nm^2\bar{\lambda}^2)$.

Independent sets in general graphs

Now for the bad news.

Given a graph G, we may efficiently construct a graph G' such that a *typical* independent set in G' points out a *maximum* independent set in G.

This constitutes a reduction from *optimisation* to *sampling*.

Theorem

There is no efficient sampler for independent sets in a general graph unless $\mathrm{RP} = \mathrm{NP}$.

Independent sets in bounded degree graphs

Restrict attention to graphs with degree bound Δ.

Independent sets in bounded degree graphs

Restrict attention to graphs with degree bound Δ.

- If Δ is sufficiently large, no efficient sampler exists unless RP $=$ NP [Luby & Vigoda]. $\Delta = 25$ suffices [Dyer, Frieze & J.]. These results use the theory of PCPs.

Independent sets in bounded degree graphs

Restrict attention to graphs with degree bound Δ.

- If Δ is sufficiently large, no efficient sampler exists unless RP $=$ NP [Luby & Vigoda]. $\Delta = 25$ suffices [Dyer, Frieze & J.]. These results use the theory of PCPs.
- If $\Delta \geq 6$ then MCMC is ineffective [DFJ].

Independent sets in bounded degree graphs

Restrict attention to graphs with degree bound Δ.

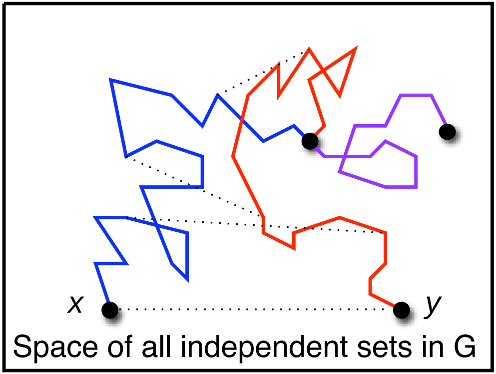
- If Δ is sufficiently large, no efficient sampler exists unless RP $=$ NP [Luby & Vigoda]. $\Delta = 25$ suffices [Dyer, Frieze & J.]. These results use the theory of PCPs.
- If $\Delta \geq 6$ then MCMC is ineffective [DFJ].
- A new algorithm makes $\Delta = 5$ tractable [Weitz, 2006].

Independent sets in bounded degree graphs

Restrict attention to graphs with degree bound Δ.

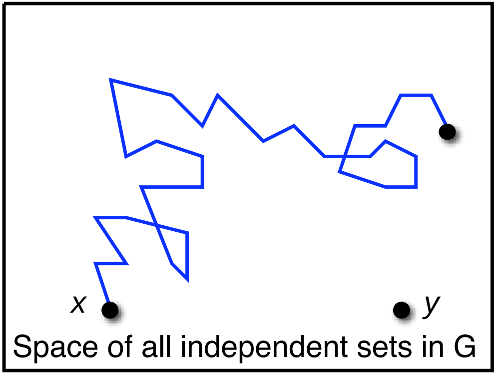
- If Δ is sufficiently large, no efficient sampler exists unless RP = NP [Luby & Vigoda]. $\Delta = 25$ suffices [Dyer, Frieze & J.]. These results use the theory of PCPs.
- If $\Delta \geq 6$ then MCMC is ineffective [DFJ].
- A new algorithm makes $\Delta = 5$ tractable [Weitz, 2006].
- $\Delta = 4$ is amenable to classical MCMC [LV].

Rough guide to coupling



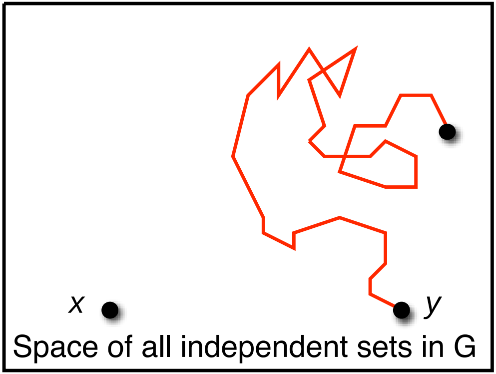
Two "coupled" evolutions of the Markov chain on the same sample space, but with different initial states.

Rough guide to coupling



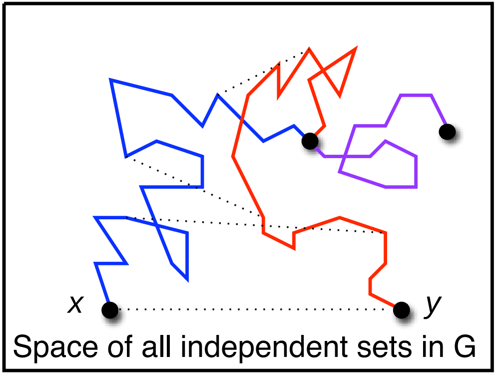
Projecting on the blue component we see a faithful copy. . .

Rough guide to coupling



Ditto projecting on red.

Rough guide to coupling



If the two can be made to coalesce rapidly, then the Markov chain must be rapidly mixing.

Independent sets in bipartite graphs: a mysterious intermediate case

The optimisation problem (find a maximum independent set in a *bipartite* graph) is in P, by network flow. So the reduction mentioned earlier does not have any complexity-theoretic consequences.

However, [Dyer, Goldberg, Greenhill & J., 2000] showed that sampling independent sets in a bipartite graph is inter-reducible with several other sampling problems (e.g., sampling downsets in a partial order). These problems are also complete for some logically defined complexity class.

A class of sampling problems of intermediate computational complexity or an illusion?

A logically defined complexity class

The complexity class containing “Bipartite Independent Set” and its peers is characterised by syntactically restricted sentences in first order logic.
E.g., the set of downsets in a partial order $(A, \prec)$ may be expressed as

$$\{D : \forall x, y \in A.\ \neg D(x) \vee \neg(y \prec x) \vee D(y)\}.$$

A logically defined complexity class

The complexity class containing “Bipartite Independent Set” and its peers is characterised by syntactically restricted sentences in first order logic.
E.g., the set of downsets in a partial order $(A, \prec)$ may be expressed as

$$\{D : \forall x, y \in A.\ \neg D(x) \vee \neg(y \prec x) \vee D(y)\}.$$

First order universal quantification.

A logically defined complexity class

The complexity class containing “Bipartite Independent Set” and its peers is characterised by syntactically restricted sentences in first order logic.
E.g., the set of downsets in a partial order $(A, \prec)$ may be expressed as

$$\{D : \forall x, y \in A.\ \neg D(x) \vee \neg(y \prec x) \vee D(y)\}.$$

CNF. (Only one clause!)

A logically defined complexity class

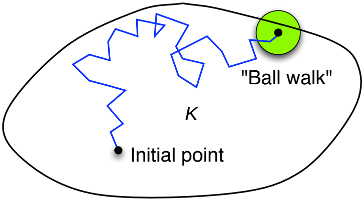
The complexity class containing “Bipartite Independent Set” and its peers is characterised by syntactically restricted sentences in first order logic.
E.g., the set of downsets in a partial order $(A, \prec)$ may be expressed as

$$\big\{D : \forall x, y \in A.\ \neg D(x) \vee \neg(y \prec x) \vee D(y)\big\}.$$

Each clause has at most one unnegated relation symbol and at most one negated relation symbol.

Highlight: sampling from a convex body

[Dyer, Frieze & Kannan, 1991], [Lovász & Simonovits, 1997].



Poincaré inequality:

$$\int_K (\nabla f(x))^2 \, dx \geq C \int_K f(x)^2 \, dx, \quad \text{for all } f \text{ with } \int_K f(x)\, dx = 0.$$

where the constant C is large if K is not "long and thin".

Some other sucesses

- Satisfying assignments to a DNF Boolean formula [Karp, Luby and Madras, 1989].
- Proper colourings of a bounded degree graph, a.k.a. antiferromagnetic Potts model. [. . . Jalsenius, Pedersen, 2006].
- Linear extensions of a partial order. [Khachiyan and Karzanov], [Bubley and Dyer].
- Feasible solutions to an instance of the knapsack problem [Morris and Sinclair].
- Perfect matchings in a bipartite graph [J., Sinclair and Vigoda].

A selection of open problems

- Is there a polynomial-time algorithm for sampling perfect matchings in a *general* graph?
- Is there an algorithm for sampling perfect matchings in a bipartite graph that is efficient in practice?
- What is the status of sampling independent sets in a bipartite graph? Is it really intermediate in complexity between independent sets in general graphs (hard for NP) and matchings in general graphs (polynomial time)?
- We are familiar with the empirical observation that "natural" decision problems tend to be in P or to be NP-complete. Is there a similar dichotomy for sampling problems? Or is there a more complex landscape, as hinted at by [Kelk, 2003]?

I. K. Brunel (9th April 1806 - 15th Sept. 1859)