NewD irections in Probability Theory

Better Coupling, Less Effort

Tom Hayes, UC Berkeley
Eric Vigoda, Georgia Tech

I: Burning Questions

General Setting

Ω: finite state space (exponential in n)

π: a distribution over Ω

Goal: Approximately sample from π in time poly(n)

Related problems: approximate counting, estimating partition functions.

Typical Setting

Ω: finite state space (exponential in n)

π: a distribution over Ω

Have a simple ergodic Markov chain with
stationary distribution π

Goal: prove "mixing time" is poly(n)

Mixing Time

$X_0, X_1, \ldots, X_t, \ldots$ distributed as the M.c.

Chain has **mixed** when $\lVert X_t - \pi \rVert_{TV} < \frac{1}{4}$.

Mixing time = $\min \{t : \forall X_0, \lVert X_t - \pi \rVert_{TV} < \frac{1}{4}\}$

Mixing Time

$X_0, X_1, ..., X_t, ...$ distributed as the M.c.

Chain has **mixed** when $\lVert \mathcal{L} X_t - \pi \rVert_{TV} < \frac{1}{4}$.

Mixing time = min $\{t : \forall X_0, \lVert \mathcal{L} X_t - \pi \rVert_{TV} < \frac{1}{4}\}$

Notes: can't replace $\forall$ with a.a.

Perhaps we can only find very weird X_0's

II: Techniques
Nice, Ice-Cold, and Old

Coupling

A technique for proving fast mixing

Easy, well-known (Doeblin, 1937)

Constructive, explicit

- Inductively matches up t-step distributions from different starts.
- Exact sampling: "Coupling from the Past"

General (at least in principle)

Often yields intuitive proofs

Coupling

For all X_0, Y_0 specify joint distribution of
$$X_1, ..., X_t, ..., Y_1, ..., Y_t, ...$$
so that, separately, (X_t) and (Y_t) are the given M.c.

Design to "coalesce": once $X_t = Y_t$, then for all $t' > t$, $X_{t'} = Y_{t'}$

Mixtime $\min \{t : \forall X_0, Y_0 \Pr(X_t \neq Y_t) < \frac{1}{4}\}$

One-Step Construction

For all X_0, Y_0 specify joint distribution of X_1, Y_1 consistent with the M.c.

Inductively evolve (X_1, Y_1) à (X_2, Y_2) à ... using the same rule.

Prove $E(\rho(X_1, Y_1)|X_0, Y_0) < (1-\varepsilon)\, \rho(X_0, Y_0)$, where ρ is integer-valued metric.

Conclude bound on mixing time.

Path Coupling Construction

Suppose ρ is a *path metric.*

For *edges* (X_0, Y_0) specify distribution of (X_1, Y_1) consistent with the M.c.

Prove $E(\rho(X_1, Y_1)|X_0, Y_0) < 1-\varepsilon$.

"Composition along paths" yields explicit one-step coupling.

Conclude bound on mixing time.

III: Hot Trail

Fearing the Worst

In coupling, must prove $\forall X_0, Y_0$

$$E(\rho(X_1, Y_1)|X_0, Y_0) < (1-\varepsilon)\, \rho(X_0, Y_0)$$

As before, $\forall$ cannot be replaced by a.a.

Burn Away the Worst

In coupling, must prove $\forall X_0, Y_0$

$$E(\rho(X_1, Y_1)|X_0, Y_0) < (1-\varepsilon)\, \rho(X_0, Y_0)$$

As before, $\forall$ cannot be replaced by a.a.

Dyer&Frieze STOC'01: suppose $\forall X_0, Y_0$

a.a. $E(\rho(X_{B+1}, Y_{B+1})|X_B, Y_B) < (1-\varepsilon)\, \rho(X_B, Y_B)$

where B is a burn-in period.

Conclude bound on coupling time.

Comments

Dyer&Frieze's method produces sharper results than traditional coupling can.

Requires clever and careful analysis of the M.c.

So far, applied very successfully to graph colorings, but not much else.

Compatible with Path Coupling (with extra care)

IV: Cool New Stuff

Coupling w/ Stationarity

Suppose we can prove a.a. X_0, $\forall\, Y_0$

$$E(\rho(X_1,Y_1)|X_0,Y_0) < (1-\varepsilon)\, \rho(X_0,Y_0)$$

Conclude essentially same bound on mixing time as for traditional coupling.

Idea: Fix Y_0, sample $X_0 \sim \pi$. Upper bound $E(\rho(X_t,Y_t)|Y_0)$ inductively rather than traditional $E(\rho(X_t,Y_t)|X_0,Y_0)$

Coupling w/ Stationarity

Need to prove a.a. X_0, $\forall Y_0$

$$E(\rho(X_1,Y_1)|X_0,Y_0) < (1-\varepsilon)\,\rho(X_0,Y_0)$$

Now we can sometimes replace burn-in argument with analysis of π

Not compatible with Path Coupling: path from typical X_0 to arbitrary Y_0 is **not** mostly typical states.

Example

For graph colorings X_0, Y_0

$$E(\rho(X_1, Y_1)|X_0, Y_0) < (1-\varepsilon)\,\rho(X_0, Y_0)$$

assuming all disagreeing verts have $> \Delta$ colors available in X_0.

Easy to check that random coloring of triangle-free graph has this property.

Improves & simplifies "Girth 5" result of Hayes (STOC '03).

Coupling w/ Stationarity II

Suppose we can only prove a.a.X_0, a.a.Y_0

$$E(\rho(X_1,Y_1)|X_0,Y_0) < (1-\varepsilon)\,\rho(X_0,Y_0)$$

Can we bound mixing time?

No! M.c. may not even be connected!

So what? Has a giant component, may mix rapidly there.

Coupling w/ Stationarity II

Suppose we can prove a.a.X_0, a.a.Y_0

$$E(\rho(X_1,Y_1)|X_0,Y_0) < (1-\varepsilon)\,\rho(X_0,Y_0)$$

We can show that if $X_0{\sim}\mu$, where μ is a "warm start" to π ($\forall X \quad \mu(X) < 2\pi(X)$), then $\lVert X_t - \pi \rVert_{TV} < (1-\varepsilon)^t \max \rho + o(1)$

Can use simulated annealing to get warm starts algorithmically. Hybrid algorithm for sampling. Not arbitrarily boostable.

Limits of 1-step coupling

Sometimes the M.c. mixes rapidly, but for every 1-step coupling, **cannot** prove

a.a.X_0, a.a.Y_0

$$E(\rho(X_1,Y_1)|X_0,Y_0) < (1-\varepsilon)\, \rho(X_0,Y_0)$$

Example: graph colorings (HV FOCS '03)

However, there is always a t-step coupling, where t is the mixing time.

Hard to construct & analyze (see HV'03)

Coupling w/ Stopping Time

Suppose we can prove $\forall X_0, Y_0$

$$E(\rho(X_T, Y_T) | X_0, Y_0) < (1-\varepsilon) \, \rho(X_0, Y_0),$$

where T is a **stopping time**, i.e., can be computed as a function of

$$X_0, \ldots, X_T, Y_0, \ldots, Y_T.$$

Can conclude a bound on mixing time.

In fact, more is true...

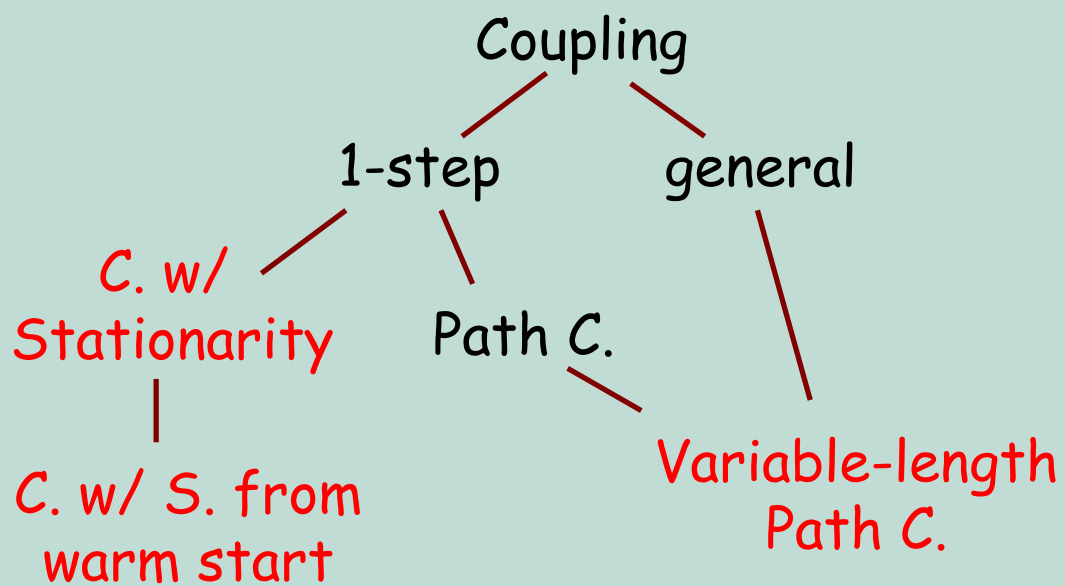# Path Coupling w/ Stop. Time

Let ρ be a *path metric.*

For *edges* (X_0, Y_0) specify distribution of $T, X_1, \ldots, X_T, Y_1, \ldots, Y_T$ consistent with the M.c., where T is a stopping time.

Prove $E(\rho(X_T, Y_T)|X_0, Y_0) < 1-\varepsilon$.

Conclude rapid mixing. (HV SODA '04).

Improved analysis possible even for one-step couplings (e.g. graph colorings for constant-degree graphs).

Overview

Coupling

1-step   general

C. w/
Stationarity   Path C.

C. w/ S. from   Variable-length
warm start   Path C.

Open Questions

Is this the final picture?

Applications besides Glauber dynamics?

Variable-length coupling from the past?

Composing variable-length partial couplings

For arbitrary colorings (X_0, Y_0), consider their shortest path
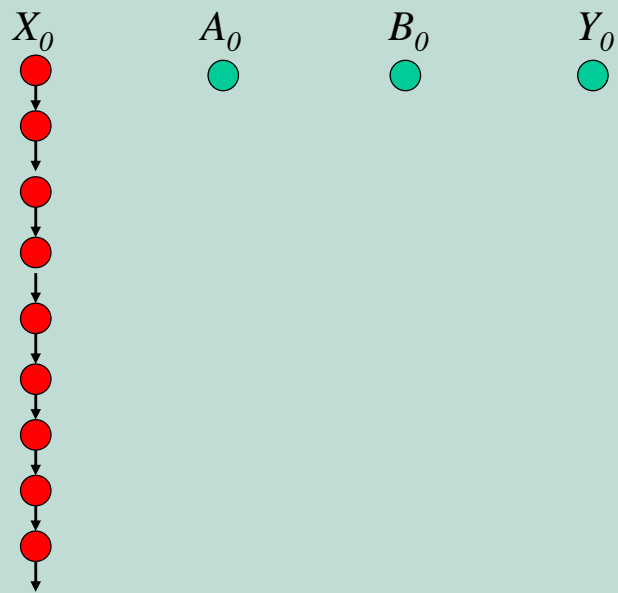
say $(Z_0, \ldots, Z_j)$
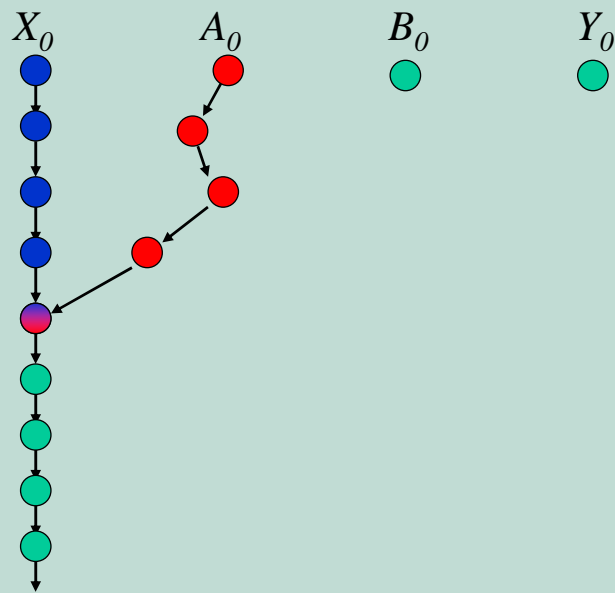
Z_0　　　　Z_1　　　　Z_2　　　　Z_3

Composing variable-length partial couplings
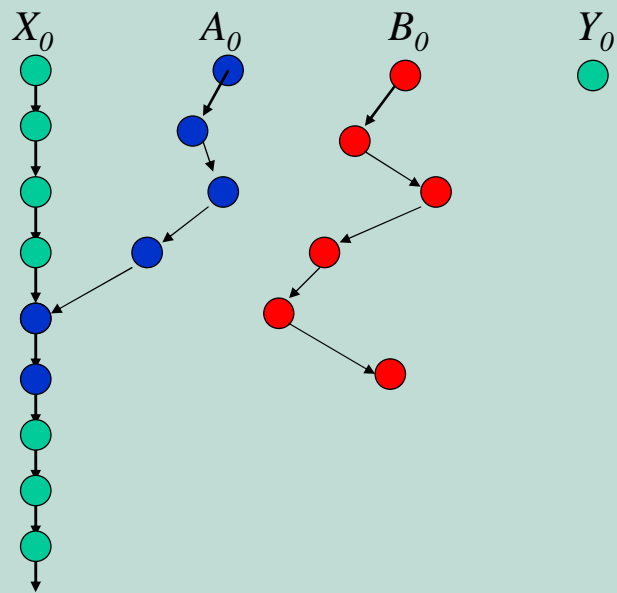
Generate $X_1, X_2, \ldots$

X_0 A_0 B_0 Y_0

Composing variable-length partial couplings
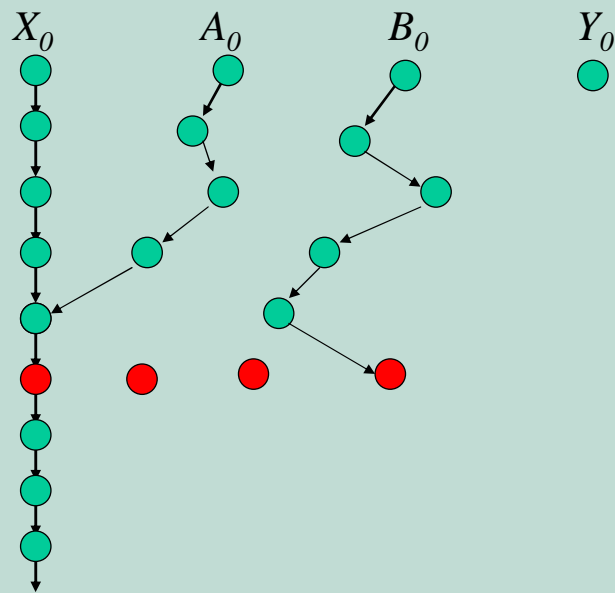
Use $X_1, X_2, ..., X_T$ to generate $A_1, ..., A_T$

Composing variable-length partial couplings

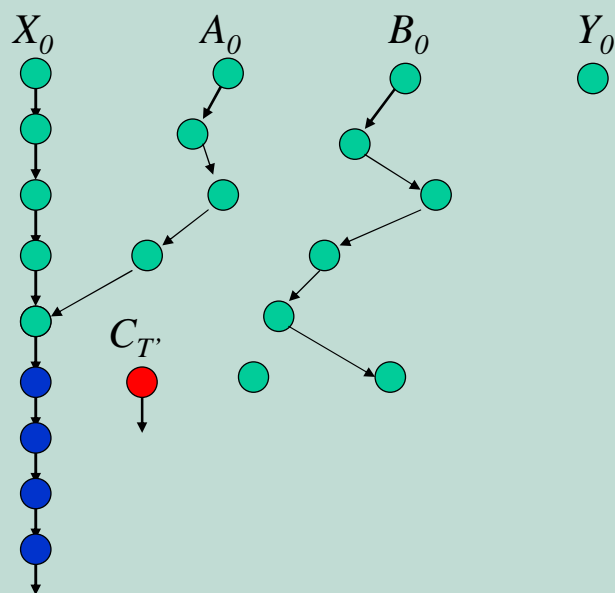*If $X_T = A_T$, use $A_1, \ldots, A_T, X_{T+1}, \ldots, X_{T'}$ to generate $B_1, \ldots, B_{T'}$*

X_0 A_0 B_0 Y_0

Composing variable-length partial couplings

If $A_{T'} \neq B_{T'}$ add the shortest path between $X_{T'}$ and $B_{T'}$

X_0 A_0 B_0 Y_0

Composing variable-length partial couplings

Recurse, i.e., use $X_{T'+1},...$ to generate $C_{T'+1},...$

X_0 A_0 B_0 Y_0

$C_{T'}$

Composing variable-length partial couplings

Recursion almost-surely terminates if $E(\rho(X_T, Y_T)) < 1$.

X_0 A_0 B_0 Y_0

$C_{T'}$
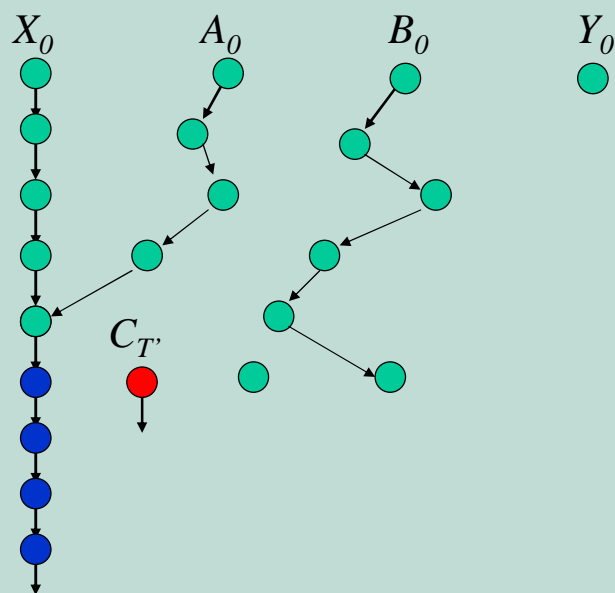
Earlier Example

For graph colorings X_0, Y_0

$$E(\rho(X_1, Y_1) | X_0, Y_0) < (1-\varepsilon)\, \rho(X_0, Y_0)$$

assuming all disagreeing verts have $> \Delta$ colors available in X_0.

Easy to check that random coloring of triangle-free graph has this property.

What's *AvailColors(z)* in a random k-coloring of a triangle-free graph with max degree Δ?

What's *AvailColors(z)* in a random k-coloring of a triangle-free graph with max degree Δ?

Given a random k-coloring and a vertex z.
Fix the coloring $\mathcal{F}$ on $V \setminus N(z)$.
Simultaneously rechoose the colors for all $w \in N(z)$.

What's *|AvailColors(z)|* in a random k-coloring of a triangle-free graph with max degree Δ?

Given a random k-coloring and a vertex z.
Fix the coloring $\mathcal{F}$ on $V \setminus N(z)$.
Simultaneously rechoose the colors for all $w \in N(z)$.

For a vertex w and color c, let

$$I(w, c) = \begin{cases} 1 & \text{if } w \text{ is colored } c \\ 0 & \text{otherwise} \end{cases}$$

Look at

$$|A(z)| = \sum_{c} \prod_{w \in N(z)} (1 - I(w, c)).$$

What's *|AvailColors(z)|* in a random k-coloring of a triangle-free graph with max degree Δ?

Given a random k-coloring and a vertex z.
Fix the coloring $\mathcal{F}$ on $V \setminus N(z)$.
Simultaneously rechoose the colors for all $w \in N(z)$.

Independent colors…

$$\mathbf{E}(A(z)|\mathcal{F}) = \sum_{c} \prod_{w \in N(z)} \left(1 - \mathbf{E}(I(w,c)|\mathcal{F})\right)$$

$$\geq k \prod_{c} \prod_{w \in N(z)} \left(1 - \mathbf{E}(I(w,c)|\mathcal{F})\right)^{1/k}$$

Chernoff!

$$= k \prod_{w \in N(z)} \left(1 - \frac{1}{A(w)}\right)^{A(w)/k}$$

w.h.p., A(z) $\approx k \exp(-\Delta/k)$

What's *|AvailColors(z)|* in a random k-coloring of a triangle-free graph with max degree Δ?

For $\Delta = \Omega(\log n), k \geq \Delta + 2$, with probability $\geq 1 - 1/n^4$, for all vertices z,

$$|A(z)| > k(\exp(-\Delta/k) - \epsilon)$$