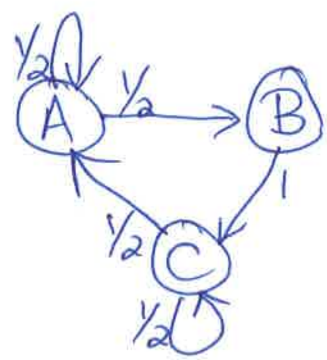
For an ergodic MC defined by P on Ω with stationary distribution π.

What if we don't know the mixing time?
Can we detect when we've reached π?

Coupling from the past by [Propp-Wilson '96].

Consider this example with 3 states $\Omega = \{A, B, C\}$



Idea: Make 3 chains X_t, Y_t, Z_t where:

$$X_0 = A,\ Y_0 = B,\ Z_0 = C.$$

Define a "global" coupling, i.e., a joint evolution of all $|\Omega|$ chains.

Let T be the 1st time that $X_T = Y_T = Z_T$.

Is this state from distribution π?
i.e., is $X_T \sim \pi$?

In this example: No it's not from π.

Why? Note $X_T \neq B$.

because we can't 1st couple at state B.

If $X_T = Y_T = Z_T = B$ then $X_{T-1} = Y_{T-1} = Z_{T-1} = A$.

But $\pi(B) > 0$ so thus X_T is not from π.

It turns out that if we do this experiment "backwards" it works.

Let's formalize:

<u>Global coupling:</u> for all states $i \in \Omega$, given the "random choice" it defines the move $i \to j$.

Think of a transition as choosing $r \in [0,1]$ uniformly at random, then the move $i \to j$ is determined by r.

Example: Ising model: $\Omega = \{+1,-1\}^V$, $V = \{0,1,\dots,n-1\}$

From $X_t \in V$,

1. Choose $r \in [0,1]$ at random.
2. If $\frac{i}{n} < r \leq \frac{(i+1)}{n}$ then:
   a. for all $j \neq i$, set $X_{t+1}(j) = X_t(j)$.
   b. Set $X_{t+1}(i) = \begin{cases} +1 & \text{if } n\left(r-\frac{i}{n}\right) < \frac{e^{\beta p}}{e^{\beta p}+e^{\beta q}} \\ -1 & \text{if } n\left(r-\frac{i}{n}\right) \geq \uparrow \end{cases}$

where p = # of + neighbors of i in X_t
& q = # of − ″

This is an equivalent form of the Glauber Dynamics/Gibbs sampler.

Transitions are defined by a function:

$$f : \Omega \times [0,1] \rightarrow \Omega$$

as long as this respects the transition matrix, i.e.,

$$\Pr\left(f(i,r) = j\right) = P(i,j)$$

Probability is over $r \in [0,1]$.

This also defines a global coupling where all chains use the same random seed $r \in [0,1]$.

1. Choose $r \in [0,1]$ at random.
2. If $X_t = i$, set $X_{t+1} = f(i,r)$.

For each time t, choose $r_t \in [0,1]$ at random.
The transitions for time $t \to t+1$ are defined by r_t.

For convenience, let $f_t : \Omega \to \Omega$ be:

$$f_t = f(\cdot, r_t)$$

i.e., $f_t(i) = f(i, r_t)$.

For a chain (X_t) we have that: $X_{t+1} = f_t(X_t)$.

Let $X_0 = i$. What's X_t?

Let
$$F_0^t(i) = f_{t-1}(f_{t-2}(\cdots f_0(i))\cdots)$$
$$= (f_{t-1} \circ f_{t-2} \circ \cdots \circ f_0)(i)$$

Then,
$$X_t = F_0^t(i)$$

More generally, for $0 \le t_1 < t_2$, let

$$F_{t_1}^{t_2}(i) = (f_{t_2-1} \circ f_{t_2-2} \circ \cdots \circ f_{t_1})(i)$$

Hence, for $i,j \in \Omega$,

$$\Pr(F_{t_1}^{t_2}(i) = j) = P^{t_2-t_1}(i,j).$$

Take the wrong forward algorithm which finds the 1st T where all chains couple.

We have N chains where $N = |\Omega|$.
Denote as $(X_t^1), (X_t^2), \dots, (X_t^N)$
where $X_0^i = i$ so ith chain starts @ $i \in \Omega$.

Run all N chains using f.
Stop when reach time T where $|F_0^T(\Omega)| = 1$
i.e., for all $i \in \Omega$, $F_0^T(i) = j$
there exists $j \in \Omega$,
(all chains are in state j at time T)

Then we output j & hope this is from π?

The earlier example shows that this is false.

What if we go back in time?

Let M be the first time where

$$|F_{-M}^{0}(\Omega)| = 1.$$

Then output $F_{-M}^{0}(\Omega)$.

<u>Theorem:</u> $F_{-M}^{0}(\Omega)$ has the same distribution as π.

<u>Proof:</u>

For fixed $t > 0$ & $i, j \in \Omega$, note:

$$\Pr(F_0^t(i) = j) = \Pr(F_{-t}^{0}(i) = j)$$

Why?

This is over $r_0, \ldots, r_{t-1}$

this is over $r_{-t}, \ldots, r_{-1}$

but same distributions, just dif't. names for those random seeds.

Thus, for all $i,j \in \Omega$,

$$\lim_{t\to\infty} \Pr\left(F_{-t}^{0}(i)=j\right) = \lim_{t\to\infty} \Pr\left(F_{0}^{t}(i)=j\right)$$

$$= \pi(j)$$ since it's ergodic with stationary dist. π

For $t = t_1 + t_2$, note:

$$F_{-t}^{0} = F_{-t_2}^{-t_1-1} \circ F_{-t_1}^{0}$$

Thus if F_{-m}^{0} is a constant function $\left(|F_{-m}^{0}(\Omega)|=1\right)$ then for all $t > m$, all $i \in \Omega$,

$$F_{-t}^{0}(i) = \left(F_{-t}^{-m-1} \circ F_{-m}^{0}\right)(i) = F_{-m}^{0}(i).$$

Hence, $F_{-m}^{0}(i) \sim \lim_{t\to\infty} F_{-t}^{0}(i) \sim \pi$ $\boxtimes$

Intuition?

Going forward: Let $T = 1^{st}$ time where F_0^t is a constant function

for $t > T$ we know F_0^t is also a constant function

not always true

but we don't know that: $F_0^t(i) \stackrel{?}{=} F_0^T(i)$

So we don't know the distribution of i.

Going backwards: for $t > M$:

$$F_{-t}^0(i) = F_{-M}^0(i)$$

So it converges to π.

because $\lim_{t \to \infty} F_{-t}^0(i) \sim \pi$.

Can we implement it efficiently?

$|\Omega|$ is HUGE so can't run $|\Omega|$ chains.

If MC is monotone then we can do it.

Ising model:

For $X_t, Y_t \in \Omega = \{+1, -1\}^V$

Say $X_t \geq Y_t$ if for all $v \in V$,

$X_t(v) \geq Y_t(v)$,

(i.e., if $Y_t(v) = +1$ then $X_t(v) = +1$)

If $X_t \to X_{t+1}$ & $Y_t \to Y_{t+1}$ use same $r \in [0,1]$

then if $X_t \geq Y_t$ then $X_{t+1} \geq Y_{t+1}$.

Why? Both chains update the same vertex.

If update v then

of + neighbors of v in X_t $\geq$ # of + neighbors of v in Y_t

hence

if $Y_{t+1}(v) = +1$ then $X_{t+1}(v) = +1$.

Consider $W_0 = \text{all} +$ & $Y_0 = \text{all} -$

& arbitrary $X_0 \in \Omega$.

Then, $W_0 \geq X_0 \geq Y_0$

& for all $t \geq 0$,

$$W_t \geq X_t \geq Y_t$$

Take t where $W_t = Y_t$ then

for all X_0, $W_t = X_t = Y_t$

So $|F_0^t(\Omega)| = 1$.

Hence, run W_t & Y_t and

look for min M so that $W_{-M} = Z_{-M}$

& then we have that $|F_{-M}^0(\Omega)| = 1$

So we're done.

So for monotone system just need to consider 2 chains, instead of $|\Omega|$ chains.

What is M?

Is it much larger than the mixing time?

For monotone MC's,

$$E[M] \leq 2T_{mix}\ln(4n).$$

So it's not much more than the mixing time.