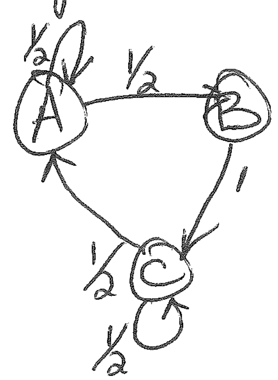
For an ergodic MC defined by P on Ω
with stationary distribution π,
- what if we don't know the mixing time?
  - can we detect when we've reached π?

<u>Coupling from the Past</u> [Propp-Wilson '96].

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



Is there a coupling that converges after T steps
& then ~~we~~ can we output that "coupled" state?

<u>Idea</u>: create 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 3 chains (in general, all 152 chains)

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

Is this state from distribution π?

i.e., is $X_T = Y_T = Z_T \sim \pi$?

In this example: NO!

Why? Note $X_T \neq B$

because we can't 1^{st} couple at B,

since if $X_T = Y_T = Z_T = B$

then $X_{T-1} = Y_{T-1} = Z_{T-1} = A$.

But $\pi(B) > 0$ so $X_T \not\sim \pi$.

Turns out we can do this approach back in time & it does work.

Let's formalize.

<u>Global coupling:</u> for all states $i \in \Omega$, given the "random choice" it defines the transition $X_t \to X_{t+1}$, for all X_t.

— think of a transition as choosing $r \in [0,1]$ uar & the transition is a function of r.

<u>Example:</u> Ising model: $\Omega = \{+1, -1\}^V$, $V = \{0, 1, .., n-1\}$.

<u>Glauber dynamics:</u>

From $X_t \in \Omega$,

1. Choose $i \in V$ uar.
2. For all $j \neq i$, set $X_{t+1}(j) = X_t(j)$.
3. Let p = # of + neighbors of i in X_t & n = # of − "

$$\text{Set } X_{t+1}(i) = \begin{cases} + & \text{w. prob. } \dfrac{e^{2\beta p}}{e^{2\beta p} + e^{2\beta n}} \\ - & \text{w. prob. } \dfrac{e^{2\beta n}}{e^{2\beta p} + e^{2\beta n}} \end{cases}$$

Global coupling:

From $X_t \in \Omega$,

1. Choose $r \in [0,1]$ u.a.r.
2. Let $i \in V$ such that: $\frac{i}{n} < r \leq \frac{(i+1)}{n}$.
3. Set $X_{t+1}(j) = X_t(j)$ for all $j \neq i$.
4. Let $r' = n\left(r - \frac{i}{n}\right)$. Note $r' \in [0,1]$.
5. Set $X_{t+1}(i) = \begin{cases} + & \text{if } r' \leq \frac{e^{2\beta p}}{e^{2\beta p} + e^{2\beta n}} \\ - & \text{o/w} \end{cases}$

Transitions are defined by a function:

$$f : \Omega \times [0,1] \rightarrow \Omega$$
$$f(X_t, r) = X_{t+1}$$

need that this respects P so that:

$$\Pr_r(f(i,r) = j) = P(i,j) \quad \text{for all } i,j \in \Omega$$

↑ Prob. over choice of $r \in [0,1]$ u.a.r.

To get a global coupling from f

Use the same r for all chains.

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

for each time t, choose random $r_t \in [0,1]$ & all transitions at time t are defined by r_t.

Let $f_t : \Omega \to \Omega$ be $f_t = f(\cdot, r_t)$

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

& then for chain (X_t) we have: $X_{t+1} = f_t(X_t)$.

Let $i = X_0$. What's X_t in terms of f?

Let $$F_0^t(i) = f_{t-1}(f_{t-2}(f_{t-3}(\cdots f_0(i))))$$
$$= (f_{t-1} \circ f_{t-2} \circ \cdots \circ f_0)(i)$$

then $X_t = F_0^t(i)$

More generally, for $0 \leq t_1 < t_2$, 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 $i, j \in \Omega$,

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

Incorrect forward algorithm:

Can rephrase as the 1^{st} T where all chains couple.

We have N chains where $N = |\Omega|$.

Denote as $(X_t^1), (X_t^2), \ldots, (X_t^N)$

where $X_0^i = i$ so i^{th} chain starts at $i \in \Omega$,

Run all N chains using f.

Stop when reach T where $|F_0^T(\Omega)| = 1$

i.e., there exists $j \in \Omega$,

and for all $i \in \Omega$, $F_0^T(i) = j$

Then we output j & "hope" $j \sim \pi$?

But we know this is not always true from the earlier example on $\{A, B, C\}$.

Go back in time?

Let M be the 1^{st} 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, i.e., $F_{-M}^{0}(\Omega) \sim \pi$.

<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 $r_0, \ldots, r_{t-1}$ — over $r_{-t}, \ldots, r_{-1}$

Same distributions just different names for these random seeds.

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

$$\lim_{t\to\infty} \Pr\left(F^0_{-t}(i)=j\right) = \lim_{t\to\infty} \Pr\left(F_0^{t}(i)=j\right)$$

$$= \pi(j)$$ since P is ergodic with stationary dist. π.

for $t = t_1 + t_2$, note:

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

thus if F^0_{-M} is a constant function (ie, $|F^0_{-M}(\Omega)|=1$)

then for all $t>M$, all $i\in\Omega$,

$$F^0_{-t}(i) = \left(F^{-M-1}_{-t} \circ F^0_{-M}\right)(i) = F^0_{-M}(i).$$

thus, $F^0_{-M}(i) = \lim_{t\to\infty} F^0_{-t}(i)$ and so we have:

Hence, $F^0_{-M}(i) \sim \lim_{t\to\infty} F^0_{-t}(i) \sim \pi$, ∎

Intuition?

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

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

but we don't know that $F_0^t(i) = F_0^T(i)$ for $t > T$,

So we don't know the distribution of $F_0^T(i)$

But going backwards,

for $t > M$,

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

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

Can we implement it efficiently?

$N = |\Omega|$ is HUGE.

Ising model is monotone,

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

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

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

(so if $Y_t(v) = +$ then $X_t(v) = +$

& if $X_t(v) = -$ then $Y_t(v) = -$)

Under global coupling (so use same r)

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

(b/c # of + neighbors of v in X_t $\geq$ # of + neighbors of v in Y_t)

of − in X_t $\leq$ # of − in Y_t

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

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

Consider $W_0 = \text{all} +$ & ~~Y_0~~ $Z_0 = \text{all} -$

then for any X_0, Y_0 we have:

$$W_0 \le X_0 \le Z_0 \ \& \ W_0 \le Y_0 \le Z_0$$

& for all $t \ge 0$,

$$W_t \le X_t \le Z_t \ \& \ W_t \le Y_t \le Z_t$$

So if $W_t = Z_t$

then $X_t = Y_t = W_t = Z_t$.

So suffices to run with ~~$X_0 = \text{all}$~~

$W_0 = \text{all} -$ & $Z_0 = \text{all} +$.

How does M compare to the mixing time?

For monotone MC's,

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

So just $O(\log n)$ slower.