Lecture 9

- Recall **Coupling from the past** (CFTP) to sample exactly from the stationary distn' of a MC $P(\cdot,\cdot)$:

  $t=0$; $F_t^0 \leftarrow$ identity fn';
  repeat
    sample random $f_t \sim \mathcal{F}$
    $F_{t-1}^0 \leftarrow F_t^0 \circ f_t$
    $t \leftarrow t-1$
  until F_t^0 is a constant function
  return the unique element in the range of $F_t^0(\cdot)$

  $\mathcal{F}$ → random function representation of $P(\cdot,\cdot)$, i.e. distribution over functions $f: \Omega \to \Omega$ s.t. $\Pr[f(x)=y] = P(x,y)$

- **CFTP Theorem [PW '96]**: If $\mathcal{F}$ guarantees that the coalescence time is finite w/pr 1, then the above procedure terminates w/prob 1 returning a value that is distributed according to π.

- However, applying the above procedure is, in general, inefficient since $|\Omega|$ is typically very large.

- **Monotone Settings:**

  Def: Suppose Ω is equiped w/ a partial order $\preceq$. A random fn' representation $\mathcal{F}$ of a MC is called monotone if $x \preceq y \Rightarrow \Pr[f(x) \preceq f(y)] = 1$.

  Claim: Suppose Ω is a partial order $\preceq$ w/ a unique minimal element $\bot$ and maximal element $\top$, and let $\mathcal{F}$ define a monotone grand coupling on Ω. Then the coupling time $T_{x,y}$ for any pair of states x,y is stochastically dominated by $T_{\bot,\top}$; i.e. $\Pr[T_{x,y} > t] \le \Pr[T_{\bot,\top} > t]$, $\forall t$.

Proof: Let $F_t = f_t \circ f_{t-1} \circ \dots \circ f_1$ where $(f_i)_{i=1}^{t}$ are independent samples from $\mathcal{F}$. After time t, (X,Y) moves to $(F_t(X), F_t(Y))$.

By monotonicity

$$F_t(\bot) \preceq F_t(X), F_t(Y) \preceq F_t(\top), \text{ wpr } 1$$

So if $F_t(\bot) = F_t(\top) \Rightarrow F_t(X) = F_t(Y), \forall x,y \text{ wpr } 1$ ☐

Back to Coupling From the Past.

- If Ω is partial order $\preceq$ w/ unique maximal, minimal elements $\bot$ and $\top$, modify above procedure as follows.

```
  -//-
  -//-
  -//-
until  F_t^0(⊥) = F_t^0(⊤)   ~> this check takes time O(t)
  -//-                          Hence, overall time spent is O(T^2)
```

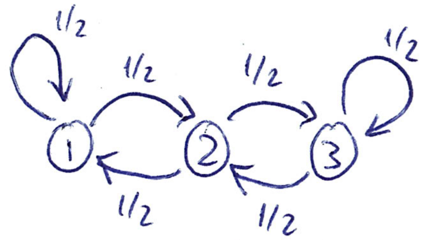
OR to the more efficient

procedure Monotone CFTP

```
T ← 1
repeat
    bottom ← ⊥
    top ← ⊤
    for t ← -T to -1 do
        bottom ← f_t(bottom)
        top ← f_t(top)
    T ← 2T
    end
until bottom = top
output top.
```

CRUCIAL: reuse same function f_t for a particular time t throughout execution; i.e. do not resample f_t for a time t that has been encountered before.

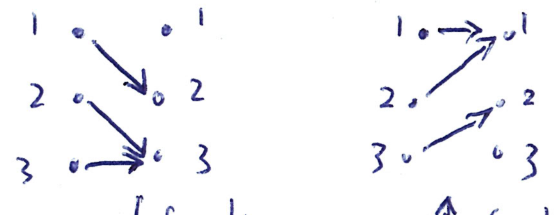
What happens if we don't re-use?



stationary distn'

$\pi(1) = \pi(2) = \pi(3)$

- consider random fn' representation:



↓ function ↑ function

F chooses ↓ w pr 1/2, ↑ w pr. 1/2

- suppose we run Monotone-CFTP procedure without re-use.
    - for $T=1$, probability of stopping is 0
    - for $T=2$: w pr $\frac{1}{4}$ stop and output 1 (case ↑,↑)

      w pr $\frac{1}{4}$ stop –//– 3 (case: ↓,↓)

      w pr $\frac{1}{2}$ continue
    - for $T=4$: w pr $\geq 1/4$ stop and output 1

      w pr. $\geq 1/4$ –//– 3

      ⋮

Hence $\Pr[\text{output } 1] \geq \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4} = \frac{3}{8} > \frac{1}{3} = \pi(1)$.

- **Running time of Monotone CFTP**

$O(T_c)$, where T_c is the coalescence time.

Bounding $\mathbb{E}(T_c)$?

Theorem [PW'96]: If Ω is partial order $\preceq$ w/ unique maximal and minimal elements $\top, \bot$ respectively.

$$\mathbb{E}(T_c) \leq O(\tau_{mix} \cdot \log h),$$

where h is the length longest chain between $\top, \bot$.

Proof: Let $X_t = F_0^t(\top)$ and $Y_t = F_0^t(\bot)$, where $F_0^t = f_{t-1} \circ f_{t-2} \circ \dots \circ f_0$, where $(f_t)_{t=0}^{+\infty}$ are iid samples from $\mathcal{F}$.

$$\mathbb{P}r[T_c > t] = \mathbb{P}r[X_t \neq Y_t]$$

$$\leq \mathbb{E}[h(X_t) - h(Y_t)] \quad \text{(where } h(\cdot) \text{ is the length of the longest chain connecting an element to } \bot\text{)}$$

$$\leq \mathbb{E}[h(X_t)] - \mathbb{E}[h(Y_t)]$$

$$\leq \|X_t - Y_t\|_{TV} \cdot h.$$

choosing $t = c \cdot \tau_{mix} \cdot \log h$ we obtain

$$\mathbb{P}r[T_c > c \cdot \tau_{mix} \cdot \log h] \leq D(c \cdot \tau_{mix} \cdot \log h) \cdot h$$

$$\leq 2 \cdot e^{-c \cdot \log h} \cdot h \leq c'.$$

$$\Rightarrow \mathbb{E}[T_c] = O(\tau_{mix} \cdot \log h). \quad \square$$

E.g. Monotone CFTP

Ising Model: Set of vertices V of a graph $G=(V,E)$ can have spin $+1$ or -1 (spin up/down respectively)

state space: $\Omega = \{+1,-1\}^V$

Gibbs Distn: $$\pi(\sigma) \propto e^{\beta(a(\sigma) - d(\sigma))}$$
$$\propto e^{2\beta a(\sigma)}$$

where $a(\sigma)$: pairs of adjacent vertices whose spins agree in σ

$d(\sigma)$: pairs of adjacent vertices whose spins disagree in σ

β: inverse temperature.

Heat-Bath MC: At state $\sigma \in \Omega$:

- pick $v \in V$ u.a.r.
- replace the spin of v by a random spin chosen according to the distribution of the spin at v in π, conditioning on the spins of v's neighbors.

i.e. set v's spin to '+' w.pr. $p_v^+ = \dfrac{e^{2\beta n_v^+}}{e^{2\beta n_v^+} + e^{2\beta n_v^-}}$

to "−" w.pr $p_v^- = 1 - p_v^+$

where $n_v^+ = |\{u \in N(v) \mid \sigma_u = +\}|$

$n_v^- = |\{u \in N(v) \mid \sigma_v = -\}|$.

<u>partial order on Ω</u>: $\sigma \preceq \tau$ iff $\sigma_v \leq \tau_v$, $\forall v \in V$

$\top$ = all spins +1

$\bot$ = all spins −1

<u>monotone grand coupling</u>:

- pick $v \in V$ u.a.r.
- pick $r \in [0,1]$ uar.
- if $r \leq P_v^+$ set spin σ_v to +1, o.w. set it to −1.

ex 0.5pt: Show this coupling is a monotone coupling.

height of partial order: $|V|$

hence $\mathbb{E}[T_c] = O(T_{mix} \cdot \log|V|)$.

Beyond Monotone Settings: The Hardcore Model.

- $G = (V, E)$, parameter $\lambda > 0$, $\Omega = \{$ independent sets of $G \} \subseteq \{0,1\}$
- $x \in \Omega$: $x_v = 1$, v is occupied

  $x_v = 0$, v is unoccupied/vacant.
- $w(x) = \lambda^{|x|}$; want to sample from $\pi(x) \propto w(x)$

<u>Heat-Bath Dynamics</u>:

- pick $v \in V$ uar. (and ignore its state)
- w pr $\frac{1}{1+\lambda}$, make v unocupied

  w pr $\frac{\lambda}{1+\lambda}$, make v occupied, if its neighbors are unoccupied & make v unoccupied, otherwise.

Exercise (1 pt): check that this MC is ergodic & reversible wrt π.

RFR:
→ pick $v \in V$ u.a.r.
→ pick $r \in [0,1]$ u.a.r.
→ use r to decide whether to try to occupy v or not

This does not define a monotone coupling.

Trick [[illegible] Häggstrom & Nelander '98, Huber '98]

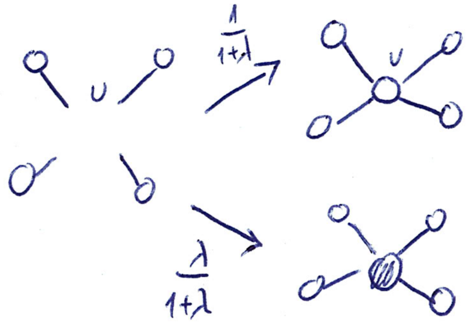
- for a set $S \subseteq \Omega$ associated a 3-valued state x

  where $x_v = 1$, if v is occupied in all states of S
  $x_v = 0$, if v is unoccupied —//—
  $x_v = ?$, if v is occupied in some states of S & unoccupied in others.

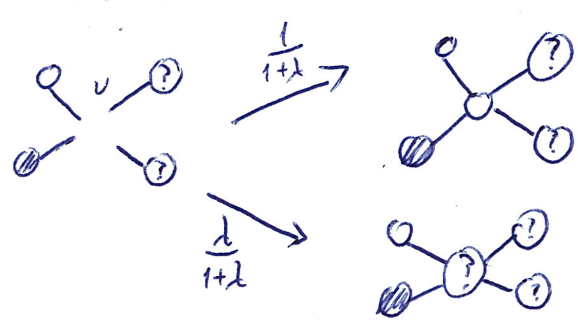
- Consider Markov Chain in ~~enlarged~~ 3-state model
  - pick $v \in V$ u.a.r.
  - with probability $\frac{1}{\lambda+1}$ propose vacancy at v (always accepted)
    w/ prob. $\frac{\lambda}{\lambda+1}$ propose occupancy (accepted only if all neighboring states are 0, o.w.)

    if fail, place "?"

case 1



case 2



RFR:
- pick v u.a.r.
- pick $r \in [0,1]$ u.a.r.
- change v's state to 0, 1, ? depending on value of r.

Ex 1pt: If in the enlarged chain, a sequence of $\{(v_t, r_t)\}_{t=1}^{T}$ converts the all-? configuration to one that has no ?, then the heat-bath dynamics of the original chain using the same sequence $\{(v_t, r_t)\}_t$ map all starting configurations to the same state.

[HN 99] Show that if $\lambda < \frac{1}{\Delta}$, Δ: maximum degree, then the # of ? decreases exponentially.

Final Remarks:
- In practice CFTP may terminate faster than the best known bound for the mixing time.
- Even if there is no known bound on T_{mix}, we can still use CFTP; when it terminates we know we have a sample from π
- We shouldn't carelessly interrupt CFTP if it takes too long & restart, as this may introduce bias in our sampling (see e.g. in page 3)
- That said, there are "interruptible" versions of CFTP see Fill '98