The Ising model

- Underlying geometry: finite graph $G = (V, E)$.
- Set of possible configurations:
  $$\Omega = \{\pm 1\}^V$$
- Probability of a configuration $\sigma \in \Omega$ given by the Gibbs distribution
  $$\mu(\sigma) = \frac{1}{Z(\beta)} \exp \left( \beta \sum_{x y \in E} \sigma(x) \sigma(y) \right)$$
  [no external field]
- Ferromagnetic $\leftrightarrow$ inverse-temperature $\beta \geq 0$.
- Goal: sample the Gibbs distribution efficiently. Main focus is on lattices at or near a certain critical $\beta_c$. 
Glauber dynamics for Ising

- One of the most commonly used MC samplers for the Gibbs distribution:
  - Update sites via iid Poisson(1) clocks
  - Each update replaces a spin at $u \in V$ by a new one $\sim \mu$ conditioned on $V \setminus \{u\}$ (heat-bath version).
- Ergodic reversible MC with stationary measure $\mu$.
- Introduced by Glauber in 1963. Other versions of the dynamics include e.g. Metropolis.
- For $\beta \geq 0$ we can couple two chains such that one is always above the other (monotone coupling).
Glauber dynamics for critical Ising

- How fast does the dynamics converge?

- 256 x 400 square lattice w. boundary conditions:
  (+) at bottom
  (−) elsewhere.

- Frame every $\sim 2^{30}$ steps, i.e. $\sim 2^{13}$ updates/site.
Example: Glauber dynamics for critical Ising on the square lattice

- 256 x 400 square lattice w. boundary conditions: (+) at bottom (-) elsewhere.
- Frame after $2^{20}$ steps, i.e. $\sim 10$ updates per site.
Strong stationary times

- Recall: Let \((X_t)\) be a Markov chain.
  - The \textit{random mapping representation} of \((X_n)\) is an i.i.d. sequence \((Z_t)\) and a map \(f\) such that \(X_t = f(X_{t-1}, Z_t)\).
  - We say that \(\tau\) is a \textit{randomized stopping time} for \((X_t)\) if it is a stopping time for such a representation \((Z_t)\).
- Def.: A \textit{strong stationary time} for a Markov chain \((X_t)\) with stationary measure \(\pi\) is a randomized stopping time \(\tau\) such that \(X_\tau \sim \pi\) independent of \(\tau\), i.e.

\[
\forall t : \mathbb{P}(\tau = t, X_\tau = y) = \mathbb{P}(\tau = t) \pi(y).
\]

\(( \iff \forall t : \mathbb{P}(\tau \leq t, X_\tau = y) = \mathbb{P}(\tau \leq t) \pi(y). \)
Bounding the mixing time

- **Theorem:** ([Aldous-Diaconis '86, '87]

  If $\tau$ is a strong stationary time for a Markov chain $(X_t)$ with stationary distribution $\pi$ then

  \[
  \max_{x \in \Omega} \left\| \mathbb{P}_x (X_t \in \cdot) - \pi \right\|_{TV} \leq \max_{x \in \Omega} \mathbb{P}_x (\tau > t).
  \]

- **Corollary:**

  Let $\tau$ be a strong stationary time for a Markov chain $(X_t)$ with stationary distribution $\pi$ and let $t_0$ be an integer such that $\max_{x \in \Omega} \mathbb{P}_x (\tau > t_0) \leq \varepsilon$. Then $t_{\text{mix}} (\varepsilon) \leq t_0$. 
Example 1: strong stationary times

- Let \((X_t)\) be a lazy simple RW on the hypercube \(\{0,1\}^n\).
- Random mapping representation: \(Z_t = (J_t, I_t)\) where \(J_t \in [n]\) and \(I_t \in \{0,1\}\) are both independent uniform.
- Strong stationary time:
  \[
  \tau_{\text{refresh}} = \min \{ t : \{J_1, \ldots, J_t\} = [n] \}. 
  \]
- By the coupon collector paradigm:
  \[
  \max_{x \in \Omega} \mathbb{P}^x \left( \tau_{\text{refresh}} > n \log n + cn \right) \leq e^{-c},
  \]
  and so
  \[
  t_{\text{mix}}(\varepsilon) \leq n \log n + \log(\frac{1}{\varepsilon}) n.
  \]
Example 2: strong stationary times

- Let \((X_t)\) be the top-to-random card shuffle: Start with \(n\) cards, repeatedly insert top into a random position.
- Strong stationary time: 1 step after bottom reaches top:
  \[
  \tau = \min \left\{ t : X_t(1) = n \right\} + 1.
  \]
- Proof: By induction, given that the cards below original bottom card (card \(\#\) \(n\)) are \(\{x_1, \ldots, x_k\}\), their ordering is uniform over \(S_k\).
- Similarly to the coupon collector:
  \[
  \tau = \tau_1 + \tau_2 + \ldots + \tau_{n-1} + 1 \quad \text{for} \quad \tau_i \sim \text{Geom}(k/n) \quad \text{ind.}
  \]
- **Corollary:**
  \[
  t_{\text{mix}}(\varepsilon) \leq n \log n + \log(\frac{1}{\varepsilon})n.
  \]
Proof (strong stationary times bound)

- Use separation distance: \( \text{sep}(t) \triangleq \max_{x, y \in \Omega} \left[ 1 - \frac{\mathbb{P}_x(X_t = y)}{\pi(y)} \right] \).

- Proof will follow from showing that:
  - Strong stationary times bound separation distance:
    \[ \text{sep}(t) \leq \max_{x \in \Omega} \mathbb{P}_x(\tau > t). \]
  - Separation distance bounds total variation distance:
    \[ \max_{x \in \Omega} \| \mathbb{P}_x(X_t \in \cdot) - \pi \|_{TV} \leq \text{sep}(t). \]
Strong stationary times bound separation distance:

If $\tau$ is a strong stationary time then for any $x, y \in \Omega$,

$$1 - \frac{\mathbb{P}_x(X_t = y)}{\pi(y)} \leq 1 - \frac{\mathbb{P}_x(X_t = y, \tau \leq t)}{\pi(y)} = \mathbb{P}_x(\tau > t)$$

and therefore $\text{sep}(t) \leq \max_{x \in \Omega} \mathbb{P}_x(\tau > t)$.

Separation distance bounds total variation distance:

$$\left\| \mathbb{P}_x(X_t \in \cdot) - \pi \right\|_{TV} = \sum_{y \in \Omega \atop \pi(y) > \mathbb{P}_x(X_t = y)} \left[ \pi(y) - \mathbb{P}_x(X_t = y) \right]$$

$$= \sum_{y \in \Omega \atop \pi(y) > \mathbb{P}_x(X_t = y)} \pi(y) \left[ 1 - \frac{\mathbb{P}_x(X_t = y)}{\pi(y)} \right]$$

and hence $\max_{x \in \Omega} \left\| \mathbb{P}_x(X_t \in \cdot) - \pi \right\|_{TV} \leq \text{sep}(t)$. 
Lower bounds on mixing

- We have seen that the top-to-random shuffle has
  \[ t_{\text{mix}}(\varepsilon) \leq n \log n + \log \left( \frac{1}{\varepsilon} \right) n. \]

  Is this tight? How do we provide lower bounds?
- Direct approach: by definition of TV distance.
- **PROPOSITION**: [Aldous-Diaconis '86]

  Let \( (X_t) \) be the top-to-random shuffle on \( n \) cards. Then for any \( \varepsilon > 0 \) there exists some \( C > 0 \) such that
  \[ d_{\text{TV}}(n \log n - Cn) > 1 - \varepsilon. \]

  In particular, \( t_{\text{mix}}(1 - \varepsilon) > n \log n - Cn. \)
Top-to-random lower bound

- Start from the inverse identity $X_0 = (n, \ldots, 1)$.
- Def: $A_j \triangleq \{\text{items } j, j-1, \ldots, 1 \text{ have original relative order}\}$
  Observe:
  - As long as card $j$ (i.e. $j$-th from bottom) did not reach the top (even +1 step) the event $A_j$ necessarily holds!
  - Stationary (uniform) probability: $\pi(A_j) = 1 / j!$
- Def: $\tau_j = \min \{t : X_t(1) = j\}$. ($j$th-bottom $\rightsquigarrow$ top)
- Proof will follow from showing that for some $C(j) > 0$,
  $$\mathbb{P}(\tau_j \geq t_n) \geq 1 - \frac{1}{\frac{j}{j-1}}$$
  where $t_n \triangleq n \log n - Cn$,
by choosing a large enough constant $j$. 
Top-to-random lower bound (ctd.)

- Remains to analyze $\tau_j$, the time it takes the $j^{th}$ from bottom card to hit the top of the deck. As before, $\tau_j$ is a sum of independent geometrics:

\[
\tau_j = \sum_{i=j}^{n-1} \tau_{j,i}, \quad \tau_{j,i} \sim \text{Geom}(i/n)
\]

\[
\mathbb{E}[\tau_{j,i}] = \frac{n}{i} \quad \text{Var}(\tau_{j,i}) < \frac{n^2}{i^2}
\]

- It follows that

\[
\mathbb{E}[\tau_j] \geq n \log n - n(1 + \log(j)),
\]

\[
\text{Var}(\tau_j) \leq \frac{n^2}{(j-1)},
\]

and Chebyshev’s inequality implies that

\[
P(\tau_j < n \log n - Cn) \leq \frac{1}{j-1}
\]

for a choice of $C = 2 + \log(j)$. \qed
Lower bounds via conductance

- Systematic approach: Relate mixing to conductance ([Lawler-Sokal ’88, Jerrum-Sinclair ’89]):
  - For a chain with transition kernel $P$ and stationary distribution $\pi$ define:
    \[ Q(x, y) \triangleq \pi(x)P(x, y) \quad ; \quad Q(A, B) \triangleq \sum_{x \in A, y \in B} Q(x, y). \]
  - The **conductance (or bottleneck ratio)** of a set $S$ is
    \[ \Phi(S) \triangleq \frac{Q(S, S^c)}{\pi(S)} \]
    and the **conductance (Cheeger constant)** of the chain is
    \[ \Phi \triangleq \min_{S : \pi(S) \leq \frac{1}{2}} \Phi(S). \]

- Intuitively: the chain is trapped inside $S$ and this represents a bottleneck for the mixing.
Examples of bottlenecks

- Binary tree on $n = 2^k - 1$ vertices:

  \[ \Phi \approx \frac{1}{n} \]
  \[ t_{\text{mix}} \approx n \]

- Two glued 2-dimensional tori on $n^2$ vertices each:

  \[ \Phi \approx \frac{1}{n^2} \]
  \[ t_{\text{mix}} \gtrsim n^2 \]
**THEOREM:**

Every Markov chain satisfies \( t_{\text{mix}} \left( \frac{1}{4} \right) \geq \frac{1}{4\Phi} \).

**PROOF:**

Let \( \mu_S \) be the stationary dist. conditioned on being in \( S \):

\[
\mu_S(x) \triangleq \frac{\pi(x) 1_{\{x \in S\}}}{\pi(S)}.
\]

By the triangle inequality

\[
\|\mu_S - \pi\|_{TV} \leq \|\mu_S - \mu_S P^t\|_{TV} + \|\mu_S P^t - \pi\|_{TV}.
\]

\[
\geq \frac{1}{2} \quad \text{due to } S^c
\]

\[
\leq \frac{1}{4} \quad \text{for } t = t_{\text{mix}}(\frac{1}{4})
\]

**CLAIM:**

This is \( \leq t \Phi(S) \)
Lower bound on mixing (ctd.)

- It remains to show \( \| \mu_S - \mu_S P^t \|_{TV} \leq t \Phi(S) \).

- Key inequality: \( \| \mu_S - \mu_S P \|_{TV} = \Phi(S) \) by definition.

- Using the fact \( \| \varphi P - \psi P \|_{TV} \leq \| \varphi - \psi \|_{TV} \) (coupling) and the triangle inequality:
  \[
  \| \mu_S P^t - \mu_S \|_{TV} \leq \| \mu_S P^t - \mu_S P^{t-1} \|_{TV} + \ldots + \| \mu_S P - \mu_S \|_{TV}
  \leq t \Phi(S).
  \]

- It now follows that \( t_{\text{mix}}(\frac{1}{4}) \Phi(S) \geq \frac{1}{4} \).
Bottlenecks in Glauber for Ising

- Recall the definition of the dynamics:
  - Update sites via iid Poisson(1) clocks
  - Each update replaces a spin at \( u \in V \) by a new one \( \sim \mu \) conditioned on \( V \setminus \{u\} \) (heat-bath version).

- How fast does it converge to equilibrium?
  - Can be exponentially slow in the size of the system:
    At low temp. (large \( \beta \)) there may be a bottleneck between “plus” and “minus” states (see tutorial).
General (believed) picture for the Glauber dynamics

- Setting: Ising model on the lattice \((\mathbb{Z}/n\mathbb{Z})^d\).
  Belief: For some critical inverse-temperature \(\beta_c\):

- Low temperature: \((\beta > \beta_c)\)
  gap\(^{-1}\) and \(t_{\text{mix}}\) are exponential in the surface area.

- Critical temperature: \((\beta = \beta_c)\)
  gap\(^{-1}\) and \(t_{\text{mix}}\) are polynomial in the surface area.

- High temperature: \((\beta < \beta_c)\)
  \- Rapid mixing: \(\text{gap}^{-1} = O(1)\) and \(t_{\text{mix}} \asymp \log n\)
  \- Mixing occurs abruptly, i.e. there is cutoff.
Gap/mixing-time evolution for Ising on the complete graph

\[
\text{gap}^{-1}, \quad t_{\text{mix}} \asymp \frac{1}{\beta - 1} \exp \left[ \frac{3}{4} (\beta - 1)^2 n \right]
\]

\[
\text{gap}^{-1}, \quad t_{\text{mix}} \asymp n^{1/2}
\]

\[
\text{gap}^{-1} = \frac{1+o(1)}{1-\beta} \quad \frac{1+o(1)}{2(1-\beta)} \log[(1-\beta)^2 n]
\]

\[
t_{\text{mix}} = O(1/\sqrt{n})
\]

(Scaling window established in [Ding, L., Peres ’09])