Mixing Time Analysis of the Glauber Dynamics for the Curie-Weiss-Potts Model

P. Cuff J. Ding E. Lubetzky Y. Peres A. Sly O. Louidor

Microsoft Research

Cornell Summer School in Probability 2009
Outline

1. Problem Definition
   - The Potts Model
   - Glauber Dynamics
   - Mixing Time
   - The Problem

2. Previous Results

3. New Results

4. Proofs
Setup and Terminology

Let $G = (V, E)$ be a finite graph, $q \in \mathbb{N}$ (number of colors) and $\hat{\beta} \in \mathbb{R}$ - inverse temperature.

A configuration $\sigma$ is an element of $\Omega = \{1, \ldots, q\}^V$.

On $\Omega$ define the (Gibbs) measure:

$$\mu(\sigma) = \mu_{\hat{\beta}, G}(\sigma) = \frac{1}{Z_{\hat{\beta}, G}} \exp \left\{ \hat{\beta} \mathcal{H}_G(\sigma) \right\}$$

where:

- $\mathcal{H}_G(\sigma) = \sum_{(u,v) \in E} \mathbb{1}_{\sigma(u) = \sigma(v)}$ - the (associated) Hamiltonian.
- $Z_{\hat{\beta}, G}$ makes $\mu$ a probability measure - the Partition Function.

Names:

- $q = 2$: The Ising Model on $G$.
- $q > 2$: The Potts Model on $G$.
- $G = C_n$ (the complete graph with $n$ vertices): Curie-Weiss Potts/Ising or Mean-field Potts/Ising.
The Question

**Question:** Given a particular sequence \((\hat{\beta}_n, G_n)_{n \geq 1}\), describe \(\mu_{\hat{\beta}_n, G_n}\) for large \(n\) or in an appropriate limit.

For Curie Weiss Potts, with \(\hat{\beta}_n = \frac{\beta}{n}\) for some \(\beta \in \mathbb{R}\), much is known (the easiest case).

For instance . . .
Fractions Vector

For $\sigma \in \Omega$ define $S(\sigma) \in S^q = \{x \in \mathbb{R}_+^q : \|x\|_1 = 1\}$ by:

$$S^k(\sigma) = \frac{1}{|V|} \sum_{v \in V} \mathbf{1}_{\{k\}}(\sigma(v)) ; \quad k = 1, \ldots, q$$

- the fractions vector.

Define $\pi_{\beta,G}$ as the distribution of $S(\sigma)$ when $\sigma$ is sampled using $\mu_{\beta,G}$.

Set $\pi_\infty = \lim_{n \to \infty} \pi_{\beta_n,G_n}$. 
Then there exists $\beta_c = \beta_c(q)$ such that:

- $\beta < \beta_c$ (high temperature):
  \[ \pi_{\frac{\beta}{n},c_n} \Rightarrow \delta_{\frac{1}{q}} \quad \text{as} \quad n \to \infty. \]
  where $\frac{1}{q}$ denotes the vector $(1/q, 1/q \ldots 1/q) \in S^q$.

- $\beta > \beta_c$ (low temperature):
  \[ \pi_{\frac{\beta}{n},c_n} \Rightarrow \sum_{k=1}^q \frac{1}{q} \delta_{T^k\hat{s}(\beta)} \quad \text{as} \quad n \to \infty. \]
  where:
  \[ \hat{s}(\beta) = \left(\hat{s}^1(\beta), \frac{1-\hat{s}^1(\beta)}{q-1}, \ldots, \frac{1-\hat{s}^1(\beta)}{q-1}\right) \in S^q \]
  $T^k$ interchanges the first and $k$-th component.

- $\beta = \beta_c$ (critical temperature):
  \[ \pi_{\frac{\beta}{n},c_n} \Rightarrow p(\beta)\delta_{\frac{1}{q}} + (1 - p(\beta)) \sum_{k=1}^q \frac{1}{q} \delta_{T^k\hat{s}(\beta)} \quad \text{as} \quad n \to \infty. \]
  where $p(\beta) \in (0, 1)$.
Phase Transition - Remarks

- \( \beta_c(q), \hat{s}(\beta) \) and \( p(\beta) \) are explicitly known.
- In fact \( \left( \pi_{\frac{p}{n}}, c_n \right)_{n \geq 1} \) satisfies a LDP on \( S^q \) with rate function:

\[
I_{\beta}(s) = R(s) - \frac{\beta}{2} \| s \| - \text{const}
\]

where \( R(s) \) is the rate function for the fractions vector when \( \beta = 0 \).

Thus, describing \( \pi_{\infty} = \pi_{\infty}(\beta) \) is solving the minimization problem of \( R(s) - \frac{\beta}{2} \| s \|_2 \) in \( S^q \).

- If \( q = 2 \), \( \hat{s}(\beta_c(2)) = \frac{1}{q} \) and the mapping \( \beta \mapsto \pi_{\infty}(\beta) \) is continuous (under the weak-topology for measures). Thus this is a second order phase transition.

- If \( q > 2 \), \( \hat{s}(\beta_c(q)) \neq \frac{1}{q} \), the mapping \( \beta \mapsto \pi_{\infty}(\beta) \) is not continuous and the phase transition is of first order. This will play a part in the rate of mixing.

- Show Graphs
Markov Chain Monte Carlo (MCMC)

A way to approximately sample from a probability measure $\mu$ on a finite space $\Omega$.

Idea: Markov Chain Monte Carlo (MCMC). Construct a Markov chain with state space $\Omega$ and $\mu$ as its stationary-ergodic distribution. Then, start from any configuration and let the chain evolve randomly for long enough time, until the distribution of the current state is close to $\mu$.

Useful when exact sampling is computationally expensive (e.g. One has to exhaust all of $\Omega$), but computing the transition probabilities is easy.

What is long enough time? One has to study the rate of convergence to stationarity - Mixing Time (later).

If $\Omega = \{1, \ldots, q\}^V$, many dynamics are possible (Glauber, Metropolis, Swendsen Wang, ...). Differ in how fast they mix.
Glauber Dynamics

Single site update dynamics for a measure $\mu$ on $\Omega = \{1, \ldots, q\}^V$:

- Start from any configuration $\sigma_0$.
- Transition:
  - Choose a vertex $u \in V$ at random.
  - Update:
    $$\sigma_{t+1}(v) = \begin{cases} 
    \sigma_t(v) & \text{if } v \neq u \\
    k & \text{if } v = u \quad \text{w.p. } \mu(\sigma(u) = k | \sigma(v) = \sigma_t(v); v \neq u)
    \end{cases}$$
- Repeat.

Conditional probabilities are straightforward if $\mu$ is a Gibbs measure (part of the definition).

$(\sigma_t)_t$ is a finite-states irreducible and aperiodic chain (at least if $\mu$ has the finite energy property), hence converges to its unique stationary distribution $\mu$.

But how fast?
Let \((X_t)_{t \in \mathbb{N}}\) be a Markov chain with state space \(S\), transition kernel \(P\) and stationary distribution \(\pi\).

Set \(d(t) = \sup_{x_0 \in S} \|P^{X_0}(X_t \in \bullet) - \pi\|_{\text{TV}}\) where \(X_0 = x_0\) under \(P^{x_0}\).
(Reminder: \(\|\mu - \nu\|_{\text{TV}} = \sup_A |\mu(A) - \nu(A)|\)).

The \(\varepsilon\)-Mixing Time of \((X_t)_{t \in \mathbb{N}}\) is:

\[
    t^M(\varepsilon) = \inf \{ t : d(t) < \varepsilon \}
\]

If no \(\varepsilon\) is specified, it is customary to use \(\varepsilon = 1/4\).
Let \((P_n)_n\) be a particular sequence of Markov chain kernels and denote by \(t^M_n(\varepsilon)\) their \(\varepsilon\)-mixing times.

We would like to know how \(t^M_n(\varepsilon)\) grows with \(n\):

- If \(t^M_n(\varepsilon)\) grows polynomially, we say that the mixing is **rapid**.
- If \(t^M_n(\varepsilon)\) grows exponentially, we say that the mixing is **slow**.

Also, if for some (and hence any) \(\varepsilon_0\) and all \(\varepsilon\):

\[
t^M_n(\varepsilon) - t^M_n(1 - \varepsilon) \triangleq w_n(\varepsilon) = o(t^M_n(\varepsilon_0)) \quad \text{as} \quad n \to \infty
\]

we say that the sequence of dynamics exhibits a **cut-off**.

- The distance to stationarity sharply changes from 1 to 0 (relatively to the mixing time).
- If \(w_n(\varepsilon) = \theta_\varepsilon(W(n))\) we say that the **cut-off window** has order \(W(n)\).
The Problem

Analyze the Mixing Time of the Glauber Dynamics for the Curie-Weiss Potts Model.

i.e., Fix $\beta$ and $q$, consider a sequence of Glauber dynamics for the Potts distribution on the $n$-complete graph: $\mu_{\beta \frac{n}{n}, C_{n}}$ and analyze the mixing time $t_{n}^{M}(\varepsilon)$ as a function of $n$. 
Previous Results

Complete analysis for the Curie-Weiss Ising case ($q = 2$):

- $\beta < \beta_c(2) = 2$ (high temperature):
  
  $t_n^M(\varepsilon) \sim \frac{1}{2} \left(1 - \frac{\beta}{2}\right)^{-1} n \log n$
  
  $w_n(\varepsilon) = \theta_{\varepsilon}(n)$

  [Aizenman, Holley ’87], [Bubley, Dyer ’97], [Levin, Luczak, Peres ’07].

- $\beta > \beta_c$ (low temperature):
  
  $t_n^M(\varepsilon)$ is exponential in $n$.

  [Griffiths, Weng and Langer ’66]

- $\beta = \beta_c$ (critical temperature):
  
  $t_n^M(\varepsilon) = \theta_{\varepsilon} \left(n^{3/2}\right)$.
  
  No cut-off.

  [Levin, Luczak, Peres ’07], [Ding, Lubetzky, Peres ’08]
Still $q = 2$ case. Now let $\beta = \beta_n$ change with $n$.

$\beta_n = \beta_c - \delta_n$:
- $\delta_n = \omega \left( \frac{1}{\sqrt{n}} \right)$ \quad $\Rightarrow$ \quad $t_n^M (\varepsilon) \sim \frac{n}{\delta} \log (\delta^2 n)$, $w_n (\varepsilon) = \theta (\frac{n}{\delta})$.
- $\delta_n = O \left( \frac{1}{\sqrt{n}} \right)$ \quad $\Rightarrow$ \quad $t_n^M (\varepsilon) = \theta (n^{3/2})$, no cut-off.

$\beta_n = \beta_c + \delta_n$:
- $\delta_n = \omega \left( \frac{1}{\sqrt{n}} \right)$ but $\delta_n = o(1)$ \quad $\Rightarrow$ \quad $t_n^M (\varepsilon) = \theta (\frac{n}{\delta} \exp \left( \left( \frac{3}{4} + o(1) \right) \delta^2 n \right))$.
- $\delta_n = \Omega (1)$ \quad $\Rightarrow$ \quad $t_n^M (\varepsilon)$ is exponential in $n$.
- $\delta_n = O \left( \frac{1}{\sqrt{n}} \right)$ \quad $\Rightarrow$ \quad $t_n^M (\varepsilon) = \theta (n^{3/2})$.
- No cut-off.

[Ding, Lubetzky, Peres ’08]
New Results - The Case $q > 2$

There exists a new critical beta $\beta_M(q) < \beta_c(q) < q$ such that:

- **$\beta < \beta_M$:**
  \[ t_n^M(\varepsilon) \sim \frac{1}{2} \left( 1 - \frac{\beta}{q} \right)^{-1} n \log n \]
  \[ w_n(\varepsilon) = \theta_{\varepsilon}(n) \]

- **$\beta > \beta_M$:**
  $t_n^M(\varepsilon)$ is exponential in $n$.

- **$\beta = \beta_M$:**
  \[ t_n^M(\varepsilon) = \theta_{\varepsilon}(n^{4/3}) \]
  No cut-off.

$\beta_M(q)$ is explicitly known.
New Results - More

Approaching criticality - $\beta = \beta_n = \beta_M - \delta_n$:

- $\delta_n = \omega \left( n^{-2/3} \right) \Rightarrow t_n^M (\varepsilon) \sim C \frac{n}{\sqrt{\delta}}, \ w_n(\varepsilon) = O_\varepsilon \left( \sqrt{\frac{n}{\delta^{5/2}}} \right)$.
- $\delta_n = O \left( n^{-2/3} \right) \Rightarrow t_n^M (\varepsilon) = \theta_\varepsilon \left( n^{4/3} \right)$, no cut-off.

Essential Mixing - $\beta_M < \beta < \beta_c$:

There exists $\Omega_n \subseteq \Omega$ with $\mu_{\beta_n, c_n} (\Omega_n) \leq e^{-cn}$ such that:

- $t_n^{M, \Omega \setminus \Omega_n} (\varepsilon) \sim \frac{1}{2} \left( 1 - \frac{\beta}{q} \right)^{-1} n \log n$
- $w_n^{\Omega \setminus \Omega_n} (\varepsilon) = \theta_\varepsilon (n)$

where $t_n^{M, \Omega \setminus \Omega_n} (\varepsilon), w_n^{\Omega \setminus \Omega_n} (\varepsilon)$ are the mixing time and cut-off window when one is not allowed to start the dynamics from $\sigma \in \Omega_n$. We say that the dynamics essentially mixes rapidly.

Low temperature - $\beta \geq \beta_c$:

$t_n^M (\varepsilon)$ is exponential in $n$. No essentially rapid mixing.
Intuition

Intuition Comes from looking at the rate function - $l_\beta(s)$.

Because of first order phase transition, near but before $\beta_c$ local minima emerge at $T^k \hat{s}(\beta)$; $k = 1, \ldots, q$.

These will slow down the mixing.

Starts to happen exactly at $\beta_M$.

Show Graphs.
Use the **Bottleneck Ratio** - For a Markov chain with transition kernel $P$ and stationary distribution $\pi$:

$$
t^M(1/4) \geq \frac{1}{4} \sup_{\substack{S \subseteq S \\ \pi(S) \leq 1/2}} \frac{\pi(S)}{\sum_{x \in S, y \notin S} \pi(x) P(x, y)} \\
\geq \frac{1}{4} \sup_{\substack{S \subseteq S \\ \pi(S) \leq 1/2}} \frac{\pi(S)}{\pi(\partial S)}
$$

Then, local minima in the rate function immediately implies exponential mixing time.
Below $\beta_M$ - Key Formula 1

Examine the fractions chain: $S_t = S(\sigma_t)$ (Markovian).

Key formula 1 - Recursion for expected distance to the equidistributed configuration:

$$\mathbb{E}\|S_{t+1} - \frac{1}{q}\|^2 = \mathbb{E}\|S_t - \frac{1}{q}\|^2 \left(1 - \frac{2\left(1 - \frac{\beta}{q}\right)}{n}\right) + \text{Error} \left(\mathbb{E}\|S_t - \frac{1}{q}\|^2, n\right)$$

$\exists \eta > 0$, such that if $\mathbb{E}\|S_0 - \frac{1}{q}\|^2 < \eta$, this gives a contraction:

$$\mathbb{E}\|S_t - \frac{1}{q}\|^2 = \left(1 - \frac{2\left(1 - \frac{\beta}{q}\right)}{n}\right)^t \|S_0 - \frac{1}{q}\|^2 + \text{Error}(n)$$
Key formula 2 - Conditional drift of one coordinate:

$$\mathbb{E} ( S_{t+1}^1 - S_t^1 \mid S_t) \leq \frac{1}{n} \left( \frac{e^{\beta S_t^1}}{e^{\beta S_t^1} + (q-1)e^{\beta(1-S_t^1)/(q-1)}} - S_t^1 \right)$$

The r.h.s is strictly negative away from $1/q$ if and only if $\beta < \beta_M$. 

Below $\beta_M$ - Key Formula 2
Below $\beta_M$ - Lower Bound on Mixing Time

Start from a configuration $\sigma_0$ with $S(\sigma_0) < \eta$

If $t_n = \frac{1}{2} \left(1 - \frac{\beta}{q}\right)^{-1} n \log n - \gamma n$, from the contraction formula:

$$\mathbb{E}\|S_{t_n} - \frac{1}{q}\|^2 \geq A(\gamma)n^{-1}$$

for $n$ large, with $A(\gamma) \to \infty$ when $\gamma \to \infty$.

By bounding the variance, this will imply that $S_{t_n}$ is far from $\frac{1}{q}$ for large $n$ with probability tending to 1 as $\gamma \to \infty$.

Since $\pi_{\frac{\beta}{n},C_n}$ concentrates around $\frac{1}{q}$, the same is true for

$$\|\mathbb{P}(S_{t_n} \in \bullet) - \pi_{\frac{\beta}{n},C_n}\|_{TV}.$$ 

Finally

$$\|\mathbb{P}(X_{t_n} \in \bullet) - \mu_{\frac{\beta}{n},C_n}\|_{TV} \geq \|\mathbb{P}(S_{t_n} \in \bullet) - \pi_{\frac{\beta}{n},C_n}\|_{TV}.$$
Start from any configuration $\sigma_0$.

Due to negative drift (Key Formula 2), after $kn$ time $\mathbb{E}\|S_{kn} - \frac{1}{q}\|^2 < \eta$.

Now we can use the contraction (Key Formula 1): If
$$t_n = kn + \frac{1}{2} \left(1 - \frac{\beta}{q}\right)^{-1} n \log n$$
Then $\mathbb{E}\|S_{t_n} - \frac{1}{q}\|^2 = O(n^{-1})$

By bounding the variance we get $\|S_{t_n} - \frac{1}{q}\| = O\left(n^{-1/2}\right)$ with as high probability as needed.

Introduce a coupling between this chain and one starting from $\pi$ such that they coincide after an additional $\gamma n$ time with probability tending to 1 as $\gamma \to \infty$. 

Use a **Coupling time** argument to bound the distance from stationarity - For any Markov chain with stationary distribution $\pi$:

$$d(t) \leq \sup_{x_0} \mathbb{P}^{x_0, \pi} (\tau_{couple} > t)$$

where $\mathbb{P}^{x_0, \pi}$ is any coupling of two copies of the Markov chain, starting from $x_0$ and $\pi$ and $\tau_{couple}$ is the first time the two processes coincide.