STAT 31512 Spring 2024 Scribe: Nathan Waniorek
April 22, 2024 Lecturer: Frederic Koehler

Cheeger's Inequality and the Random Cluster Model

1 The Last Couple of Lectures

- Dobrushin's condition (sampling under weak dependence).
- Sampling is possible even though exact counting is hard (sampling $\approx$ approximate counting).
- Tightness of Dobrushin (Curie-Weiss), where the Glauber dynamics do not mix.
- Sampling algorithms beyond Glauber.

2 Curie-Weiss Model

We will show the phase transition at $\beta = 1$ for the Curie-Weiss model. Recall that the model is defined as follows:

$$x \in \{\pm 1\}^n, \quad \pi(x) = \frac{1}{Z_\beta} \exp\left(\frac{\beta}{2n}\left(\sum_{i=1}^n x_i\right)^2\right).$$

This model is exactly solvable ("integrable"). We use the following trick: For $x \sim \pi$, define $M = \sum_{i=1}^n X_i$. The random variable M is called the "magnetization".

$$\pi(M = m) = \sum_{x : \sum x_i = m} \pi(x) = \exp\left(\frac{\beta}{2n} m^2\right) \cdot \binom{n}{\frac{n+m}{2}}.$$

Stirling's approximation: $\log(n!) = n \log n - n + \mathcal{O}(\log n)$. Recall that

$$\binom{n}{pn} = \frac{n!}{(pn)!((1-p)n)!}.$$

We compute that

$$\begin{aligned}
\log \binom{n}{pn} &= n \log n - n - (pn \log pn - pn) - ((1-p)n \log(1-p)n - (1-p)n) \pm \mathcal{O}(\log n) \\
&= n\left(p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}\right) \pm \mathcal{O}(\log n) \\
&= nH(Ber(p)) \pm \mathcal{O}(\log n).
\end{aligned}$$

Plugging this calculation into our expression for $\pi(M)$,

$$\pi(M = m) = \exp\left(\frac{\beta}{2n} m^2 + nH\left(Ber(\frac{1}{2} + \frac{m}{2n})\right) \pm \mathcal{O}(\log n)\right).$$

$$\tilde{m} = \frac{m}{n}, \quad \pi(M = m) = \exp\left(n\left(\frac{\beta}{2}\tilde{m}^2 + Ber(\frac{1}{2} + \frac{\tilde{m}}{2})\right) \pm \mathcal{O}(\log n)\right).$$

We define $f_\beta(\tilde{m}) = n\left(\frac{\beta}{2}\tilde{m}^2 + Ber(\frac{1}{2} + \frac{\tilde{m}}{2})\right)$. Taylor expanding around $\tilde{m} = 0$ gives us

$$f_\beta(\tilde{m}) = H(Ber(\frac{1}{2}) + \frac{\beta}{2}\tilde{m}^2 - \frac{1}{2}\tilde{m}^2 + \mathcal{O}\left(\tilde{m}^4\right),$$

where we used the fact that $Ber(\frac{1}{2} + \frac{\tilde{m}}{2})$ is even in $\tilde{m}$.

Fact: when $\beta \leq 1$, then $f_\beta(\tilde{m})$ is concave. By concavity, the global maxima of f_β is obtained at $\tilde{m} = 0$. Conversely, if $\beta > 1$, f_β is bimodal, with symmetric global maxima at $\tilde{m}^*$ and $-\tilde{m}^*$.

$$\pi\left(\tilde{M} \in [\tilde{m}^* - \varepsilon, \tilde{m}^* + \varepsilon]\right) \to 1$$

as $n \to \infty$ for any $\varepsilon > 0$. If $\tilde{m}$ is contained in this region, then $f_\beta(\tilde{m}) < f_\beta(\tilde{m^*} - \delta$ for some $\delta > 0$. Consequently,

$$\pi(\tilde{M} = \tilde{m}) \leq \exp\left(-\delta n \pm \mathcal{O}\left(\log n\right)\right).$$

Taking a union bound, we get that

$$\pi\left(\tilde{M} \notin [\tilde{m}^* - \varepsilon, \tilde{m}^* + \varepsilon]\right) \leq n \exp\left(-\delta n \pm \mathcal{O}\left(\log n\right)\right).$$

In the bimodal case, this means the mass is concentrated entirely at the two modes, and there is no way for the dynamics to cross between them.

We want to show that the Glauber dynamics is “torpidly mixing”. Rigorously, we want to show that $1 - \lambda_2(P) = \exp\left(-\Theta(n)\right)$ when $\beta > 1$. Recall the Poincaré inequality:

$$\mathrm{Var}(f) \leq \frac{1}{1 - \lambda 2} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}(f|x_{\sim i})\right] = \frac{1}{1 - \lambda_2} \mathcal{E}_\pi(f, f).$$

We see that it suffices to find one “bad” f where $\mathrm{Var}(f) \gg \mathcal{E}(f, f)$. We consider the following f:

$$f(x) = f(M) = \begin{cases} +1 & \text{if } M \geq 1, \\ -1 & \text{if } M \leq -1, \\ M & \text{otherwise.} \end{cases}$$

One can verify that $\mathrm{Var}(f) = 1$. However, if $\beta > 1$ we have that $\mathbb{E}_{x_{\sim i}}[\mathrm{Var}(f(x)|x_{\sim i})] = \exp\left(-\Theta(n)\right)$, because $\mathbb{E}_{x_{\sim i}}[\mathrm{Var}(f(x)|x_{\sim i})] = 0$ unless $\sum_{j \neq i} x_j \in [-4, 4]$ and $\pi(\sum_{j \neq i} x_j \in [-4, 4]) = \exp\left(-\Theta(n)\right)$. Consequently, we have that $1 \leq \frac{1}{1-\lambda_2} \exp\left(-\Omega(n)\right)$, completing the proof. One can also show a matching lower bound for the spectral gap. We have demonstrated a “bottleneck” for the dynamics.

3 Cheeger's Inequality

Definition 1 (Bottleneck Ratio)**.**

$$\Phi = \min_{S:\pi(S) \leq \frac{1}{2}} \frac{\mathcal{E}_p(\mathbf{1}_S, \mathbf{1}_S)}{\pi(S)} = \frac{\text{“surface area of S”}}{\text{”volume of S”}}.$$

The bottleneck ratio is a notion of isoperimetry. The following result shows that the bottleneck ratio provides upper and lower bounds on the spectral gap.

Theorem 1 (Cheeger's Inequality)**.**

$$\frac{\Phi^2}{2} \leq 1 - \lambda_2 \leq 2\Phi.$$

The upper bound is referred to as the "easy direction" and the lower boud as the "hard direction".

Proof idea: For the easy direction, given S, look at $f(x) = \mathbf{1}_{\{x \in S\}}$, use the Poincaré inequality to show that $\mathrm{Var}(f) \leq \frac{1}{1-\lambda_2} \mathcal{E}_p(\mathbf{1}_S, \mathbf{1}_S)$. For the hard direction, consider level sets $S_t = \{x : f(x) > t\}$. Then prove that if $\frac{\mathcal{E}_p(f,f)}{\mathrm{Var}(f)}$ is small, then there must exist t such that $\frac{\mathcal{E}_f(f_{S_t}, f_{S_t})}{\pi(S_t)}$ is small.

Exercise: verify that the easy direction is tight for $\{\pm 1\}^n$ and the hard direction is tight for cycles.

4 What to do when Glauber fails to mix?

e.g. Curie-Weiss with $\beta > 1$, which have shown exhibits torpid mixing.

Theorem 2 (Jerrum-Sinclair '93)**.** *For all graphs G, for any $\beta \geq 0$, letting*

$$\pi(x) = \frac{1}{Z} \exp\left(\beta \sum_{i \sim j} x_i x_j\right),$$

there exists a poly$(n, \log \frac{1}{\varepsilon})$ time sampler that achieves TV error ε.

How do we sample?

$$Z = \sum_x \prod_{i \sim j} \exp\left(\beta x_i x_j\right),$$

and note that $\beta x_i x_j = 2\beta \left[\mathbf{1}[x_i = x_j] - \frac{1}{2}\right]$.

Fact: define $p = 1 - e^{2\beta}$, then $e^{2\beta(\delta - 1)} = 1 - p + p\delta$, for $\delta \in \{0, 1\}$. Consequently,

$$Z \propto \sum_x \prod_{i \sim j} \left[1 - p + p\mathbf{1}[x_i = x_j]\right].$$

Definition 2 (Edwards-Sokal/Fortuin-Kasteleyn-Swendsen-Wang Coupling)**.** For $x \in \{\pm 1\}^n$, $y \in \{0, 1\}^E$, define

$$\mu(x, y) = \frac{1}{Z} \prod_{i \sim j} \left[(1 - p)\mathbf{1}(y_{ij} = 0) + p\mathbf{1}(x_i = x_j)\mathbf{1}(y_{ij} = 1)\right].$$

By the calculation above:

$$\mu(x) = \sum_y \mu(x, y) = \pi(x).$$

On the other hand,

$$\begin{aligned}
\mu(y) = \sum_x \mu(x, y) &\propto \sum_x \prod_{i \sim j} \left[(1 - p)\mathbf{1}(y_{ij} = 0) + p\mathbf{1}(y_{ij} = 1)\mathbf{1}(x_i = x_j)\right] \\
&\propto p^{\#\{y_{ij}=1\}}(1 - p)^{\#\{y_{ij}=0\}} \sum_x \mathbf{1}\left(x_i = x_j \quad \forall i, j \text{ s.t. } y_{ij} = 1\right) \\
&\propto p^{\#\{y_{ij}=1\}}(1 - p)^{\#\{y_{ij}=0\}} 2^{\#\text{connected components in } y}.
\end{aligned}$$

This is known as the "Fortuin-Kasteleyn random cluster model".

Fact:

$$\mu(x|y) = \frac{\mu(x,y)}{\mu(y)} = \frac{\mathbf{1}(x \text{ satisfies } y)}{2^{\#\text{connected components in } y}}.$$

Following from the above calculations:

$$\mu(y|x) \propto \prod_{i \sim j : x_i = x_j} \left[(1-p)\mathbf{1}(y_i j = 0) + p\mathbf{1} y_{ij} = 1)\right] \prod_{ij : x \neq x_j} \left[\mathbf{1}(y_{ij} = 0)\right],$$

which is a product measure. This is known as the "Bond percolation" on $\{i \sim j : x_i = x_j\}$.

Definition 3 (Swendsen-Wang Dynamics)**.** The Swendsen-Wang dynamics are defined as follows:

1. Initialize at some $x_0 \in \{\pm 1\}^n$.
2. For $t = 1 \dots T$:
   (a) Sample $y^{(t)} \sim \mu(y = \cdot | x = x^{(t-1)})$
   (b) Sample $x^{(t)} \sim \mu(x = \cdot | y = y^{(t)})$

Theorem 3 (Guo-Jerrum '16)**.** *The Swensen-Wang dynamics have* $1 - \lambda_2 = \Omega\left(\frac{1}{poly(n)}\right)$.

In fact, Guo and Jerrum prove this theorem by proving the following stronger result:

Theorem 4. *The Gibbs sampler for* y *has* $1 - \lambda_2 = \frac{1}{poly(n)}$.

They then reduce the previous theorem to this one.