CS294 Markov Chain Monte Carlo: Foundations & Applications **Fall 2009**

Lecture 19: November 10

Lecturer: Prof. Alistair Sinclair *Scribes: Kevin Dick and Tanya Gordeeva*

19.1 Independent Sets (Continuation)

In the previous lecture, we discussed the problem of sampling independent sets uniformly at random.

Input: Graph $G = (V, E)$ on $|V| = n$ vertices.

Goal: Sample an independent set of vertices of G uniformly at random.

Definition 19.1 *Call a Markov Chain with state space Ω equal to the independent sets of G and uniform stationary distribution γ-*cautious *if, in any step, it changes the disposition of at most γn vertices of G.*

We'll prove the following theorem, due to Dyer, Frieze, and Jerrum [DFJ]:

Theorem 19.2 *There exist $\gamma > 0$ and graph G with maximum degree $\Delta = 6$ such that any γ-cautious Markov chain on independent sets of G has mixing time* $\exp(\Omega(n))$.

Note that this result is best possible: a theorem of Weitz [Wei] says that there exists a local Markov chain (which changes one vertex per step) with mixing time $O(n \log n)$ for all G with $\Delta \leq 5$.

To prove the theorem, we'll use a result from the previous lecture:

Theorem 19.3 *For any Markov chain, and for all $S \subseteq \Omega$ with $\pi(S) \leq 1/2$, $\tau_{mix} \geq \frac{1}{4\Phi(S)}$, where*

$$\Phi(S) = \frac{C(S, \overline{S})}{\pi(S)} = \sum_{x \in S, y \in \overline{S}} \frac{\pi(x)P(x,y)}{\pi(S)}.$$

Corollary 19.4 $\tau_{mix} \geq \frac{|S|}{4|\partial S|}$, *where ∂S is the* inner boundary *of S, i.e., the set of states in S that are connected to some state outside S.*

Proof: (of Theorem 19.2) Let $G = (L, R, E)$ be a *random* bipartite graph, with $|L| = |R| = n$, constructed by throwing down Δ independent random perfect matchings (and erasing duplicate edges). Clearly the maximum degree is at most Δ.

Definition 19.5 *An (α, β)-*independent set *is one with αn vertices in L and βn vertices in R.*

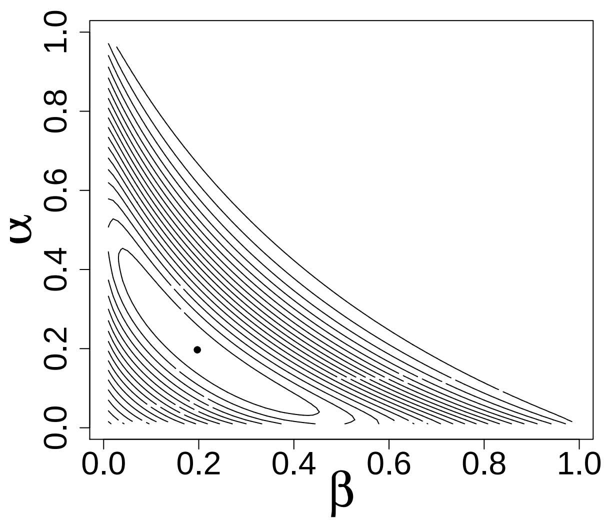

Figure 19.1: $f(\alpha, \beta)$ with $\Delta = 5$.

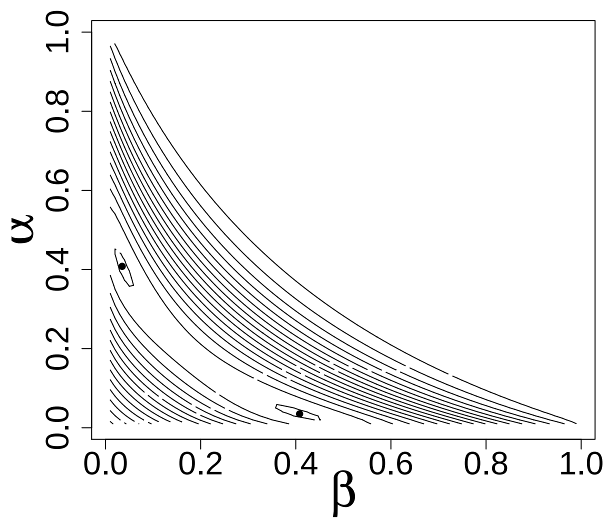

Figure 19.2: $f(\alpha, \beta)$ with $\Delta = 6$.

Let $\mathcal{E}(\alpha, \beta)$ be the expected number of (α, β)-independent sets. We have $\binom{n}{\alpha n}$ ways to choose αn vertices in L and $\binom{n}{\beta n}$ ways to choose βn vertices in R. Moreover, the probability that a given set of αn vertices in L won't hit some given set of βn vertices in R under a single random perfect matching is the probability that the αn outgoing edges from L only hit the other $(1-\beta)n$ vertices in R, which is $\binom{(1-\beta)n}{\alpha n} / \binom{n}{\alpha n}$. Since we have Δ independent random perfect matchings, we see that

$$\mathcal{E}(\alpha, \beta) = \binom{n}{\alpha n}\binom{n}{\beta n}\left[\binom{(1-\beta)n}{\alpha n} \Big/ \binom{n}{\alpha n}\right]^{\Delta}.$$

By Stirling's approximation, $m! \sim \sqrt{2\pi m}(m/e)^m$, we have

$$\binom{n}{\alpha n} = \left(\frac{1}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}\right)^n \theta\left(\frac{1}{\sqrt{n}}\right),$$

which yields

$$\mathcal{E}(\alpha, \beta) = \left[\frac{(1-\beta)^{(\Delta-1)(1-\beta)}(1-\alpha)^{(\Delta-1)(1-\alpha)}}{\alpha^{\alpha}\beta^{\beta}(1-\alpha-\beta)^{\Delta(1-\alpha-\beta)}}\right]^{n+o(n)}.$$

Thus we can write $\mathcal{E}(\alpha, \beta) = \exp\{f(\alpha, \beta)(n + o(n))\}$, where

$$f(\alpha, \beta) := -\alpha \ln \alpha - \beta \ln \beta - \Delta(1-\alpha-\beta)\ln(1-\alpha-\beta) + (\Delta-1)((1-\alpha)\ln(1-\alpha) + (1-\beta)\ln(1-\beta)).$$

The behavior of f is qualitatively very different for the cases of $\Delta \leq 5$ and $\Delta \geq 6$. In particular, note the following properties:

- f is symmetric in α, β, and has no local maxima (other than its global maxima).
- When $\Delta \leq 5$, there is a unique maximum, which falls on the line $\alpha = \beta$ (Figure 19.1).
- When $\Delta \geq 6$, there are two (symmetric) maxima, neither of which fall on the line $\alpha = \beta$ (Figure 19.2).

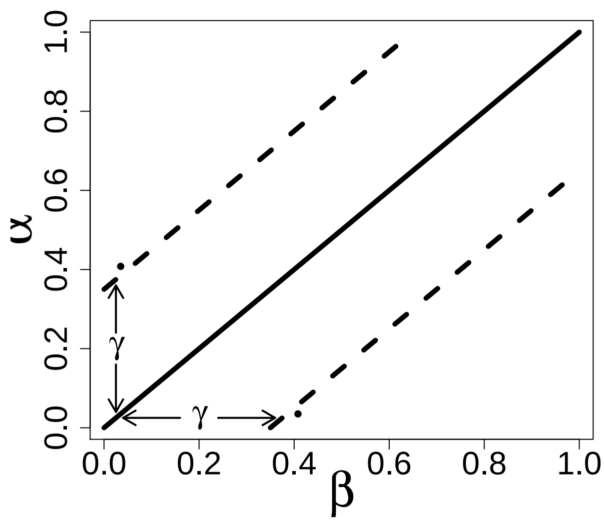

Figure 19.3: Strip of width 2γ between the two maxima.

We'll study the case of $\Delta \geq 6$. Let (α^*, β^*) be the location of (one of the two) maxima. (For $\Delta = 6$, $\alpha^* \approx 0.035$ and $\beta^* \approx 0.408$.) Numerically, $f(\alpha^*, \beta^*) > c = 0.71$. So $\mathcal{E}(\alpha^*, \beta^*) \geq e^{cn}$ for sufficiently large n.

Choose $\gamma = .35$. Then, again numerically, the maximum value of $f(\alpha, \beta)$ on the strip of width 2γ in Figure 19.3 is less than $c - \delta$, for some $\delta \geq .0001$. Since there are clearly at most n^2 values of (α, β) in this strip, the expected number of independent sets with (α, β) in this strip is at most $n^2 e^{(c-\delta)n} < e^{c'n}$, for some $c' < c$ and sufficiently large n.

Let I_{left} be the set of (α, β)-independent sets with $\alpha > \beta$, I_{right} the set of (α, β)-independent sets with $\alpha < \beta$, and I_{mid} the set of (α, β)-independent sets with (α, β) in the strip. Ignoring the possible case of $\alpha = \beta$, we see that $I_{\text{mid}} \subseteq I_{\text{left}} \cup I_{\text{right}}$.

By Markov's inequality, $|I_{\text{mid}}| \leq e^{c'n}$ with high probability (with $c' = 0.706$). Deterministically, the number of (α, β)-independent sets I with $|I \cap L| = \alpha n$ is at least $\binom{n}{\alpha n} 2^{(1-\Delta\alpha)n}$, since for every choice of αn vertices in L we can select some subset of the (at least) $(1 - \Delta\alpha)n$ vertices in R that aren't adjacent to any of the chosen vertices in L. By a Stirling-like argument as above, this quantity is $\exp(g(\alpha)(n - o(n)))$, where $g(\alpha) = -\alpha \ln \alpha - (1 - \alpha) \ln(1 - \alpha) + (\ln 2)(1 - \Delta\alpha)$.

Setting $\Delta = 6$ and optimizing over α gives a lower bound on $\exp(g(\alpha)(n - o(n)))$ of $e^{c''n}$, where $c'' > c'$ (in fact, $c'' = .708$).

Clearly this analysis holds for both $I \cap L$ and $I \cap R$. Choose S to be the smaller of I_{left} and I_{right}, so that $\pi(S) \leq 1/2$. Then

$$\frac{|\partial S|}{|S|} \leq \frac{e^{c'n}}{e^{c''}n} = \frac{1}{e^{(c''-c')n}} \leq e^{-0.002n}.$$

By Theorem 19.3, this completes the proof. ■

19.2 The Ising model in two dimensions

Recall that the Ising model assigns $+$ or $-$ to vertices in a $\sqrt{n} \times \sqrt{n}$ box in the two dimensional square lattice with the probability of a configuration σ given by the *Gibbs distribution*:

$$\pi(\sigma) = Z^{-1} \exp\left(\sum_{i \sim j} \beta \sigma_i \sigma_j\right) = \exp\left(\beta(\#\text{ of agreeing neighbors} - \#\text{ of disagreeing neighbors})\right).$$

(The factor Z is just a normalizing factor, or partition function.) The sum is over adjacent pairs i, j. The parameter β is the inverse of the temperature (in physics, the model simulates ferromagnetism). When β is low, the Gibbs distribution approaches the uniform distribution over configurations, corresponding to the fact that at high temperatures σ will not exhibit macroscopic order. Conversely, when β is high (low temperatures), the distribution assigns much higher weight to configurations that are highly organized. There is a critical (inverse) temperature β_{crit} at which the model undergoes a *phase transition*, that is, it switches between being organized and being disorganized. A more formal approach to β_{crit} examines the correlation between the value of the spin at the origin and the value of spins at the boundary as n tends to infinity. For $\beta < \beta_{\text{crit}}$, the spin at the origin will be uncorrelated with the boundary as $n \to \infty$; for $\beta > \beta_{\text{crit}}$, there may be positive correlation with the boundary (e.g., for the all-plus or all-minus boundary) even as $n \to \infty$.

$$\begin{array}{cc} & \sqrt{n} \\ \sqrt{n} & \boxed{\begin{array}{ccccc} + & - & - & + & - \\ - & - & + & - & - \\ + & - & + & + & + \\ - & + & - & - & - \\ - & - & - & + & - \end{array}} \end{array}$$

Figure 19.4: An example of a configuration on the 5x5 lattice

The heat bath (Glauber) dynamics is a Markov chain on spin configurations starting from any initial configuration. At each step, a site i is chosen u.a.r. from the lattice, and σ_i is resampled from the conditional distribution on σ_i given the neighbors. Let m_i^+ be the number of neighbors of i with value $+$, and let m_i^- be the number of neighbors with value $-$. Given that i is chosen, the probability that the new spin at i will be $+$ is given by $\frac{\exp(\beta(m_i^+ - m_i^-))}{\exp(\beta(m_i^+ - m_i^-)) + \exp(\beta(m_i^- - m_i^+))}$. This gives a reversible Markov chain that converges to the Gibbs distribution.

The following remarkable fact about the Glauber dynamics has been proved relatively recently for such a classical model (see, e.g., [M98]):

Theorem 19.6 *The mixing time of the Glauber dynamics for the Ising model on a $\sqrt{n} \times \sqrt{n}$ box in the 2-dimensional square lattice is:*

$$\begin{cases} O(n \log n) & \text{if } \beta < \beta_c; \\ e^{\Omega(\sqrt{n})} & \text{if } \beta > \beta_c, \end{cases}$$

where β_{crit} is the critical point (i.e., phase transition).

The full proof of this result requires more time than we have available in this class. However, we'll give elementary arguments that show that the mixing time is $O(n \log n)$ for sufficiently small β, and that it is

$\exp(\Omega(\sqrt{n}))$ for sufficiently large β. Pushing both these results to the threshold value β_{crit} is technically rather involved.

For the rest of this lecture, we concentrate on showing a lower bound $\exp(\Omega(\sqrt{n}))$ on the mixing time for $\beta > \beta_0$ for some constant β_0 (sufficiently low temperature). We can think of the configurations at low temperatures as being divided into two "phases" where the spins are mostly $+$ or mostly $-$ respectively, with a low probability of transitioning between the phases. (At high temperatures such a bottleneck does not exist because the "intermediate" configurations, which are the most numerous, have sufficiently large weight.)

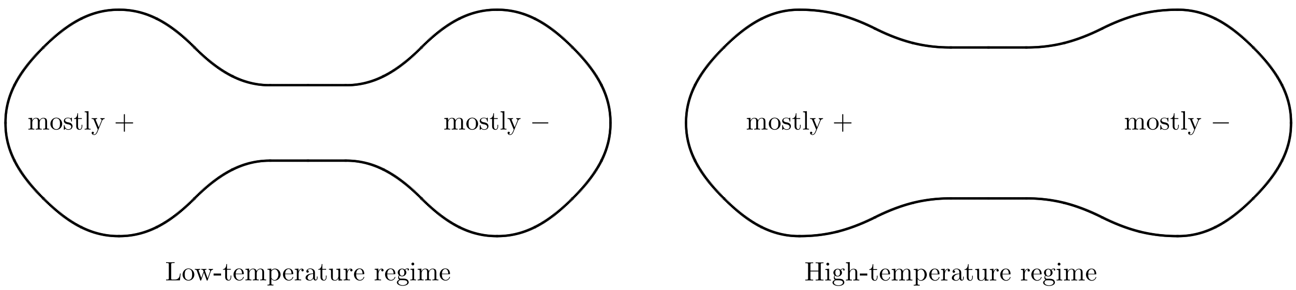

Figure 19.5: The low-temperature regime has a bottleneck between states of mostly $+$ and mostly $-$.

Claim 19.7 *There exists a large, but finite β such that the mixing time is* $\exp(\Omega(\sqrt{n}))$.

The key idea in the proof is the concept of a *fault line*, due to [Ran] and [Per].

Definition 19.8 *A* fault line *is a line that crosses the $\sqrt{n} \times \sqrt{n}$ box either left-right or top-bottom and separates $+$ from $-$.*

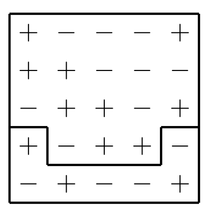

Figure 19.6: An example of a fault line

Fact 1: Let F be the set of configurations that contain a fault line. Then $\pi(F) \leq e^{-C\sqrt{n}}$ for some $C > 0$, provided β is large enough.

Proof: Fix a fault line L with length ℓ. Let $F(L)$ be the set of configurations containing ℓ. Take any configuration $\sigma \in F(L)$, and flip the spins of σ on one side of the fault line. The weight of σ will go up by a factor of $e^{2\beta\ell}$ since $\sigma_i\sigma_j$ only changes for i, j neighbors on opposite sides of the fault line. For a fixed fault line L, this operation is one-to-one on $F(L)$. Therefore $\pi(F(L)) < e^{-2\beta\ell}$. Now sum over L to get

$$\pi(F) \leq 2\sqrt{n} \sum_{\ell \geq \sqrt{n}} \gamma^\ell e^{-2\beta\ell} \leq e^{-c\sqrt{n}} \quad \text{when } \beta > \frac{1}{2}\ln\gamma$$

The factor of $2\sqrt{n}$ accounts for the starting points of the fault lines and whether the fault line is left-right or top-bottom. The quantity γ is the connective constant, $\gamma = \lim_{\ell\to\infty} |\Gamma_\ell|^{1/\ell}$ where $|\Gamma_\ell|$ is the number of self-avoiding walks of length ℓ. For our purposes, it's enough that $\gamma \le 3$ since there are at most 3 directions to choose from at each step in the fault line. ■

Fact 2: If there does not exist a + crossing or a − crossing from left to right, then there must be a fault line going top-bottom.

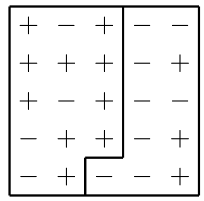

Figure 19.7: An example of a top-bottom fault line when there are no monochromatic left-right crossings

Proof: (Sketch) There must a leftmost path L of + vertices from top to bottom and a rightmost path R of − vertices from top to bottom, otherwise there would be either a − crossing or a + crossing from left to right. Suppose L is always to the right of R (otherwise take the rightmost path of +s and the leftmost path of −s). Then it is possible to construct a fault line between L and R with + to the left of the fault line and − to the right of the fault line. ■

Fact 3: If there exists a vertex v that has a + path and a − path to the top, then there exists a "fault line" from v to the top (ie, a line starting from v to the top that separates + and −).

We leave the proof of Fact 3 as an **exercise** (see Figure 19.8 for inspiration).

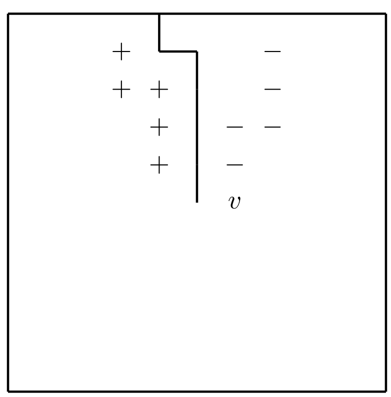

Figure 19.8: The vertex v has paths of +s and −s to the top, so it has a fault line to the top.

Now define S_+ to be the set of configurations with a left-right + crossing and a top-bottom + crossing. Define S_- similarly, except with − crossings. If $\sigma \in \overline{(S_t \cup S_i)}$, then σ either has no monochromatic left-right crossing or no monochromatic top-bottom crossing; hence it has a fault line by Fact 2. Therefore $\pi(S_-) = \pi(S_+) \to 1/2$ as $n \to \infty$ by Fact 1.

Let $\tilde{\delta} S_+$ be the *exterior boundary* of S_+ (i.e., configurations adjacent to S_+ that are not in S_+).

Claim 19.9 $\pi(\tilde{\delta} S_+) \le e^{-c'n}$.

Proof: If $\sigma \in \overline{(S_+ \cup S_-)}$, then σ has a fault line and therefore has exponentially small probability by Fact 1. So consider $\sigma \in \tilde{\delta} S_+ \cap S_-$. For any such σ there must be a vertex v with spin $-$ having $+$ crossings from v to the top, bottom, left, and right. (This follows because flipping the spin at v must put the configuration in S_+.) There also must be a $-$ crossing from left to right, or from top to bottom, that passes through v (because flipping v takes the configuration out of S_-); w.l.o.g. suppose there exists a left-right such crossing. Therefore there are fault lines from the left to v and from v to the right. Concatenating these two fault lines gives us a true top-bottom fault line, except that the two lines may not match up exactly at v (e.g., one may end at the top of v and the other at the bottom of v). Hence we get a fault line with at most a small (constant) number of imperfections. But it should be clear that the proof of Fact 1 is robust enough to handle a constant number of imperfections. So once again, σ has exponentially small probability. ■

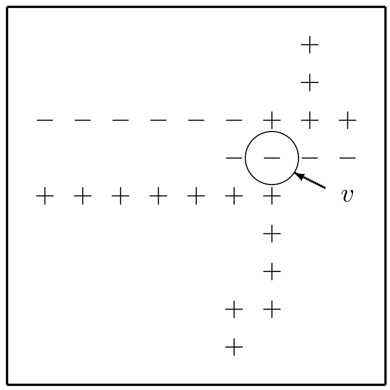

Figure 19.9: There is one vertex which blocks top-bottom and left-right + crossings.

References

[DFJ] M. Dyer, A. Frieze and M. Jerrum. "On counting independent sets in sparse graphs." *SIAM Journal on Computing* **31** (2002), pp. 1527–1541.

[M98] F. Martinelli. Lectures on Glauber dynamics for discrete spin models. *Lectures on Probability Theory and Statistics*, Springer Lecture Notes in Mathematics **1717**, 1998, pp. 93–191.

[Per] Y. Peres. "On the lower bound for mixing time of Glauber dynamics for the Ising model." Manuscript, 2006.

[Ran] D. Randall. "Slow mixing of Glauber dynamics via topological obstructions." *Proceedings of SODA* 2006, pp. 870–879.

[Wei] D. Weitz. "Counting independent sets up to the tree threshold." *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*, 2006, pp. 140–149.