Approximate Counting and Mixing Time of Markov chains **Fall 2024**

Lecture 8: An FPRAS for the Ferromagnetic Ising Model

Lecturer: Shayan Oveis Gharan *Oct 23*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

In this lecture, we will present a classic and fundamental result due to Jerrum and Sinclair[JS93] that shows how to estimate the partition function of the ferromagnetic Ising model using the path technology.

These notes are heavily cribbed from notes written by Alistair Sinclair.[1]

Definition 1.1 (The Ferromagnetic Ising Model)**.** *Let $G = (V, E)$ be an undirected graph, let $\beta \geq 0$ (the "inverse temperature"), and let h be any real number (the "external field") Associated with each vertex $i \in V$ is a* ***spin*** *in $\sigma_i \in \{-1, +1\}$; the set of spins constitutes a* ***configuration*** *$\sigma \in \pm 1^n$. The* ***weight*** *of the configuration is defined to be*

$$w(\sigma) := \exp\left(\sum_{(i,j)\in E} \beta\sigma_i\sigma_j + \beta h \sum_{i\in V} \sigma_i\right).$$

The ***Gibbs distribution*** *associates probability*

$$\pi(\sigma) = \frac{w(\sigma)}{Z}$$

to configuration σ, where

$$Z := Z(G, h, \beta) = \sum_{\sigma\in\pm 1^n} w(\sigma)$$

is the ***partition function*** *of the ferromagnetic Ising model.*

Our goal is to estimate the partition function Z, a quantity of fundamental importance in statistical physics, and specifically, to prove the following theorem. This gives us another example of how to use the flow encoding technique from the previous lecture.

Theorem 1.2. *[JS93] There is an FPRAS for the ferromagnatic Ising partition function.*

As we've seen for other problems, the first thing we might naturally try is to design a Markov chain on configurations σ whose stationary distribution is π, show that this Markov chain is rapidly mixing, and then use self-reducibility or some other technique to estimate the partition function given approximately random samples.

Unfortunately, the natural Markov chain,i.e., the Glauber dynamics, in which we flip one spin at a time is not rapidly mixing. For example, if β is very large and $h = 0$, then the distribution is concentrated on configurations with almost all spins the same, either nearly all 1 or nearly all -1. As a consequence, it can be shown that the time it takes to get from the all +1 configuration to the all -1 configuration using these local dynamics is exponentially large. Therefore, at least if we want to devise a chain that works for any ferromagnetic parameters, we will have to take a different route.

[1] Lecture 13 in CS294 at Berkeley from Fall 2023.

1.0.1 Reduction to the subgraphs world

The approach will be to use a classical result [NM53] to develop a connection between the Ising model and a seemingly unrelated model called the "subgraphs world", whose configurations are subgraphs of G, and then show how to use samples from this other distribution to estimate Z. This

Definition 1.3 (The subgraphs world)**.** *Let $0 < \eta, \xi \leq 1$, $G = (V, E)$ and let $\Omega = \{A : A \subseteq E\}$. Define*

$$\hat{w}(A) = \xi^{|odd(A)|}\eta^{|A|},$$

where $odd(A)$ is the set of vertices of odd degree in $H = (V, A)$ and let

$$\hat{\pi}(A) = \frac{\hat{w}(A)}{\hat{Z}},$$

where

$$\hat{Z} := \hat{Z}(\eta, \xi) = \sum_{A \subseteq E} \eta^{|A|}\xi^{|odd(A)|}$$

is the partition function of the subgraphs world.

The next theorem shows that, somewhat surprisingly perhaps, up to an easily computable constant factor, Z and $\hat{Z}$ are the same and hence we can estimate Z using an estimate for $\hat{Z}$.

Theorem 1.4. *Given the Ising model with graph $G = (V, E)$ (with $|V| = n$) and parameters β and h, let $\eta := \tanh(\beta)$, $\xi := \tanh(\beta h)$ and $C_{\beta,h} = \cosh(\beta)^{|E|}\cosh(\beta h)^{|V|}$. Then*

$$Z(G, \beta, h) = C_{\beta,h)} \cdot \hat{Z}(\eta, \xi)$$

Proof. Recalling that for any configuration of the Ising model, we have $w(\sigma) := \exp\left(\sum_{(i,j)\in E} \beta\sigma_i\sigma_j + \beta h \sum_{i \in V} \sigma_i\right)$, we apply the identity

$$e^x = \cosh(x)(1 + \tanh(x))$$

to rewrite $w(\sigma)$ as follows:

$$w(\sigma) = \prod_{(i,j)\in E} \cosh(\beta\sigma_i\sigma_j) \prod_{(i,j)\in E} (1 + \tanh(\beta\sigma_i\sigma_j)) \prod_{k\in V} \cosh(\beta h\sigma_k) \prod_{k \in V}(1 + \tanh(\beta h \sigma_k))$$

which, since cosh is an even function and tanh is an odd function, equals

$$= \cosh(\beta)^{|E|}\cosh(\beta h)^{|V|} \prod_{(i,j)\in E} (1 + \sigma_i\sigma_j \tanh(\beta)) \prod_{k \in V}(1 + \sigma_k \tanh(\beta h))$$

$$= C\left(\sum_{A\subseteq E} \tanh(\beta)^{|A|} \prod_{(i,j)\in A} \sigma_i\sigma_j\right)\left(\sum_{S \subseteq V} \tanh(\beta h)^{|S|} \prod_{i \in S}\sigma_i\right)$$

where $C = \cosh(\beta)^{|E|}\cosh(\beta h)^{|V|}$

$$= C \sum_{A\subseteq E, S\subseteq V} \tanh(\beta)^{|A|}\tanh(\beta h)^{|S|} \prod_{i \in \text{odd}(A)} \sigma_i \prod_{j\in S}\sigma_j \qquad \text{since } \sigma_k^2 = 1$$

$$= C \sum_{A\subseteq E, S\subseteq V} \tanh(\beta)^{|A|}\tanh(\beta h)^{|S|} \prod_{i \in \text{odd}(A)\oplus S} \sigma_i$$

Now, define $\eta := \tanh(\beta)$ and $\xi := \tanh(\beta h)$. Then

$$\begin{aligned}\hat{Z} &= \sum_{\sigma \in \{\pm 1\}^n} C \sum_{A \subseteq E, S \subseteq V} \eta^{|A|} \xi^{|S|} \prod_{i \in \mathrm{odd}(A) \oplus S} \sigma_i \\ &= C \sum_{A \subseteq E, S \subseteq V} \eta^{|A|} \xi^{|S|} \sum_{\sigma \in \{\pm 1\}^n} \prod_{i \in \mathrm{odd}(A) \oplus S} \sigma_i \\ &= C \sum_{A \subseteq E} \eta^{|A|} \xi^{|\mathrm{odd}(A)|}\end{aligned} \tag{1.1}$$

since $\sum_{\sigma \in \{\pm 1\}^V} \prod_{i \in B} \sigma_i$ is equal to 2^n if $B = \emptyset$ and 0 otherwise. Thus the only choices of S that survive the inner summation in (1.1) above are those with $S = \mathrm{odd}(A)$. □

Thus, if we can estimate $\hat{Z}$, then we can estimate Z.

1.1 The algorithm for sampling from $\hat{\pi}$.

We use the Metropolis Glauber dynamics to sample from $\hat{\pi}$, where $\Omega = \{A\}_{A \subseteq E}$. We denote the state at time t by $F_t \subseteq E$. The transitions are then defined as follows:

- With probability 1/2, $F_{t+1} := F_t$.
- Otherwise, pick $e \in E$ uniformly at random and set

$$F_{t+1} := \begin{cases} F_t \oplus e & \text{with probability } \min\left(\frac{\hat{w}(F \oplus e)}{\hat{w}(F_t)}, 1\right) \\ F_t & \text{otherwise.} \end{cases}$$

It is immediate that this chain is irreducible, aperiodic and has stationary distribution $\hat{\pi}$. We will prove the following theorem.

Theorem 1.5. *For every graph $G = (V, E)$ with n vertices and m edges and every $0 < \xi, \eta \leq 1$, the Glauber dynamics, starting from the empty graph, has mixing time*

$$T_{mix} \leq O\left(\frac{m^3}{\xi^6}\right).$$

When we apply this theorem later, we will ensure that $\xi = \Omega(1/n)$.

1.2 Analysis of the Markov chain using path technology

We prove Theorem 1.5 using the path technology from the previous lecture. Let us review the basic definitions, assuming that we are analyzing a Markov chain with state space Ω, kernel $K(\cdot, \cdot)$ and stationary distribution π.

- For each pair of states $x, y \in \Omega \times \Omega$, we choose a path $P_{x \to y}$ and route $\pi(x)\pi(y)$ units of flow along that path. We will call this path the *canonical path* for the pair x, y.

- The total flow $f(t)$ on transition $t = (u, v)$ in the Markov chain is

$$\sum_{(x,y)\in\Omega^2 \text{ s.t. } u\to v\in P_{x,y}} \pi(x)\pi(y).$$

- Given the above paths, the *congestion* $c(t)$ on the transition $t = u \to v$ is defined to be

$$c(t) = \frac{f(t)}{Q(t)} := \frac{1}{\pi(u)K(u\to v)} \sum_{(x,y)\in\Omega^2 \text{ s.t. } t=u\to v\in P_{x,y}} \pi(x)\pi(y).$$

- In the previous lecture, we proved the following bound due to Jerrum and Sinclair:

$$\frac{1}{\alpha} \leq \ell_{\max} \cdot \max_{u\to v} \; c(u\to v) \tag{1.2}$$

  where α is the Poincare constant and $\ell_{\max} = \max_{x,y} |P_{x,y}|$. Plugging this bound into the standard bound on mixing time based on the Poincare' constant, we obtain

$$T_{mix}(\epsilon, s) \leq \ell_{\max} \cdot \max_{u\to v} \; c(u\to v) \left(\log(4\pi(s))^{-1} + 2\log(\epsilon^{-1})\right). \tag{1.3}$$

  where $T_{mix}(\epsilon, s)$ is the number of steps to get the total variation distance from π down to ϵ starting from state $s \in \Omega$.

Suppose, for example, that we wanted to use this technique to show that the mixing time of a Markov chain is polynomial in parameter n. In most applications, the canonical paths are shortest paths, hence bounded in length by the diameter of the Markov chain which is polynomial in n. Suppose also that π is uniform and the transition probabilities are at least $1/\text{poly}(n)$. Then we would need to show that the number of canonical paths that pass through any transition of the chain is at most $|\Omega| \cdot$ (polynomial in n). This is particularly challenging if we do not not even know what $|\Omega|$ is. To get around this, Jerrum and Sinclair came up with an ingenious approach known as ”flow encodings”.

1.2.1 Flow encodings

The idea is the following: for each transition $u \to v$ construct an injective mapping from the pairs (x, y) whose canonical paths pass through transition $u \to v$ to the state space Ω. Since the map is injective, any state $z \in \Omega$ picks out at most one path (x, y) that uses the transition $t = u \to v$, namely the (x, y) that are mapped to z by the mapping associated with transition t. Thus, z can be thought of as an ”encoding” of the canonical path. The goal is to define this injective mapping so that the flow $\pi(x)\pi(y)$ routed from x to y is roughly proportional to $\pi(z)$. This enables us to bound the total flow through each transition and hence the congestion on any transition in the Markov chain.

In more detail, we need the following definitions and lemma.

Definition 1.6 ($\mathcal{Q}_{u\to v}$)**.** *Denote by $\mathcal{Q}_{u\to v}$ the set of pairs $(x, y) \in \Omega^2$ such that path $P_{x,y}$ goes through the transition $u \to v$.*

Definition 1.7 (Flow encoding)**.** *Given the set of canonical paths $\{P_{x\to y}\}_{x,y\in\Omega^2}$, and the induced sets $\{\mathcal{Q}_{u\to v}\}$, one per transition $u \to v$ of the Markov chain, a **flow encoding** for the set of canonical paths is a collection of injective maps, one per transition of the Markov chain*

$$\{enc_t : \mathcal{Q}_t \to \Omega\}_{t=(u,v)}.$$

Lemma 1.8. *Let $\{enc_t\}$ be a* ***flow encoding*** *for the canonical paths, such that for some $\gamma > 0$, for all $x, y \in \Omega^2$ and any transition $t = u \to v$ such that $(x, y) \in \mathcal{Q}_{u \to v}$*

$$\frac{\pi(x)\pi(y)}{\pi(u)K(u \to v)} \leq \gamma \cdot \pi(enc_{u \to v}(x, y)).$$

Then the maximum congestion $\max_{u \to v}\ c(u \to v)$ is at most γ.

Proof. Consider any transition $u \to v$. Then

$$\begin{aligned} c(u \to v) &= \frac{1}{\pi(u)K(u \to v)} \sum_{x,y \in \Omega^2\ :\ (x,y) \in \mathcal{Q}_{u \to v}} \pi(x)\pi(y) \\ &\leq \gamma \cdot \sum_{x,y \in \Omega\ :\ (x,y) \in \mathcal{Q}_{u \to v}} \pi(\text{enc}_{u \to v}(x, y)) \end{aligned}$$

which by injectivity of the map $\text{enc}_{u \to v}$

$$\leq \gamma \cdot \sum_{z \in \Omega} \pi(z) \leq \gamma.$$

□

1.2.2 Canonical paths for the subgraphs model

We now apply the canonical paths/flow encoding technique to the subgraphs model. To avoid confusion and make it clear that states correspond to subsets of edges, we will name the states of the Markov Chain by capital letters. (So concretely, what we might have called x and y above, we are now calling I and F, and what we called u, v above, we are now calling A, B, where $I, F, A, B \subseteq E$.)

The first step is to define the canonical paths $P_{I,F}$ from any initial state I to any final state F.

Specifically, for a pair of states I, F (each corresponding to a subset of edges in E) let

$$\Delta = I \oplus F$$

be the symmetric difference of the set of edges in I and the set of edges in F. Suppose that Δ has $2k$ vertices of odd degree. (Necessarily there is an even number of vertices of odd degree.) Then the edges in Δ can be covered with a collection of edge-disjoint paths and circuits $C_1, \ldots, C_r$, where $C_1, \ldots, C_k$ are paths that start and end at a vertex of odd degree, and $C_{k+1}, \ldots, C_r$ are circuits.

To find these, simply pick an odd vertex in (V, Δ) and walk without repeating edges until the first time another odd vertex is encountered. Then delete all covered edges. Repeat until k paths have been found. At this point, all remaining vertices in Δ have even degree, so the remaining edges can be covered by a union of Eulerian tours. Since the paths and circuits covering the edges in Δ are not uniquely identified, use some predefined rule to pick out a particular choice, that includes a distinguished initial vertex in each path or circuit and a direction. For example, number the vertices, always start from the lowest numbered vertex and choose the neighboring edge that connects to the lowest numbered vertex.

Definition 1.9 (The canonical paths)**.** *The canonical path from I to F consists of "fixing" the paths and circuits in sequence from C_1 to C_r. To process a particular path or circuit C_i, with transitions in order $t_1, \ldots, tj$, move along those edges, adding the corresponding edge if it's in F and removing the edge if it's in I.*

Definition 1.10 (The flow encoding)**.** *For each* $(I, F) \in \mathcal{Q}_{A\to B}$*, where* $B = A \oplus e$ *is a transition of the Markov chain, define*

$$enc_{A\to B}(I, F) = I \oplus F \oplus (A \cup B).$$

Lemma 1.11. *The above flow encoding is injective. In other words, the transition* $A \to B$ *and* $z \in \Omega$ *(in the range of* $enc_{A\to B}$*) uniquely determine the pair* I, F *s.t.* $enc_{A\to B}(I, F) = z$*.*

Proof. First observe that

$$\text{enc}_{A\to B}(I, F) \oplus (A \cup B) = I \oplus F \oplus (A \cup B) \oplus (A \cup B) = I \oplus F.$$

Moreover, since $A \oplus B = e$ is the edge that is added or deleted by the transition $A \to B$, we know exactly which portions of the paths/cycles have been "processed" so far and where in the canonical path we are. We also know that A agrees with F on the paths and circuits already processed and with I on the rest from which we can uniquely reconstruct I and F. □

Lemma 1.12. *For any pair of states* I, F *and any transition* $t = A \to B$ *on the canonical path from* I *to* F*, we have*

$$\hat{w}(I)\hat{w}(F) \le \xi^{-4}\ \hat{w}(A \cup B) \cdot \hat{w}(enc_t(I, F)).$$

Proof. Recall that by definition $\hat{w}(A) = \xi^{|\text{odd}(A)|}\eta^{|A|}$, where $\text{odd}(A)$ is the set of vertices of odd degree in $H = (V, A)$. Since the multiset of edges in $I \uplus F$ and and $(A \cup B) \uplus \text{enc}_t(I, F)$ are the same, $\eta^{|I|}\eta^{|F|} = \eta^{|A\cup B|}\eta^{\text{enc}_t(I,F)}$, it suffices to show that

$$\xi^{\text{odd}(I)}\xi^{\text{odd}(F)} \le \xi^{-4}\xi^{\text{odd}(A\cup B)}\xi^{\text{odd}(\text{enc}_t(I,F))}. \tag{1.4}$$

We analyze the contribution to (1.4) from each vertex. Start with the observation that for any two sets of edges C, D,

$$v \in \text{odd}(C) \oplus v \in \text{odd}(D) \equiv v \in \text{odd}(C \oplus D). \tag{1.5}$$

Now, for a vertex v, let[2]

$$\alpha(v) = \mathbf{1}_{v\in\text{odd}(A\cup B)} + \mathbf{1}_{v\in\text{odd}(\text{enc})} - \mathbf{1}_{v\in\text{odd}(I)} - \mathbf{1}_{v\in\text{odd}(F)} \in \{-2, \dots, 2\}.$$

Let *processed* be the set of edges toggled in the prefix of the canonical path $P_{I,F}$ up to the point when the state $A \cup B$ is reached. Then

$$A \cup B = \textit{processed} \oplus I \qquad \text{and} \qquad \text{enc} = \textit{processed} \oplus F \tag{1.6}$$

Now, suppose that $v \in \text{odd}(I \oplus F)$. Then we get the following chain of implications:

$$v \in \text{odd}(I \oplus F) \underset{1.5}{\Longrightarrow} v \in \text{odd}(I) \text{ iff } v \notin \text{odd}(F) \underset{1.6}{\Longrightarrow} v \in \text{odd}(A \cup B) \text{ iff } v \notin \text{odd}(enc)$$

Thus $\alpha(v) = 0$.

Otherwise, v has even degree in $I \oplus F$. Suppose first that v also has even degree in *processed*. Then:

$$\begin{aligned} v \notin \text{odd}(I \oplus F) \quad \text{and} \quad v \notin \text{odd}(\textit{processed}) &\underset{1.6}{\Longrightarrow} v \in \text{odd}(I) \text{ iff } v \in \text{odd}(A \cup B) \\ \text{and} &\underset{1.6}{\Longrightarrow} v \in \text{odd}(F) \text{ iff } v \in \text{odd}(enc) \end{aligned}$$

[2]Note that henceforth we are writing $\text{enc} := \text{enc}_t(I, F)$.

so again $\alpha(v) = 0$.

It remains to consider the case where $v \notin \mathrm{odd}(I \oplus F)$ and $v \in \mathrm{odd}(\mathit{processed})$. But this implies that v is either the final endpoint of the last edge flipped in $A \cup B$ or the start vertex of the current circuit, if one is being processed. In either of these cases, it could be that $\alpha(v) = 2$, which altogether proves (1.4).

□

Putting it all together, we obtain the following bound:

Lemma 1.13. *The mixing time starting from the empty set is*

$$T_{mix}(\epsilon, \emptyset) = O(m^2 \xi^{-6} \log \epsilon^{-1}).$$

Proof. We need to find the γ s.t.

$$\frac{\pi(I)\pi(F)}{\pi(A)P(A \to B)} \leq \gamma \cdot \pi(\mathrm{enc}).$$

where $\mathrm{enc} = I \oplus F \oplus (A \cup B)$.

Now, we have $B = A \oplus e$ and so by reversibility of the chain,

$$\pi(A)P(A \to B) = \pi(A \cap B)P(A \cap B \to A \cup B),$$

so

$$\hat{w}(A \cap B)P(A \cap B \to A \cup B) = \hat{w}(A \cap B)\frac{1}{2m}\min\left(\frac{\hat{w}(A \cup B)}{\hat{w}(A \cap B)}, 1\right) = \frac{1}{2m}\min(\hat{w}(A \cup B), \hat{w}(A \cap B))$$

so

$$\frac{\pi(I)\pi(F)}{\pi(\mathrm{enc})\pi(A \cap B)P(A \cap B \to A \cup B)} = 2m \cdot \frac{\hat{w}(I)\hat{w}(F)}{\min(\hat{w}(A \cup B), \hat{w}(A \cap B))\hat{w}(\mathrm{enc})} \underset{\text{Lemma 1.12}}{\leq} \frac{2m}{\xi^6}$$

In the final inequality, we also used the fact that η in $[0, 1]$, and $A \cup B$ and $A \cap B$ differ on only one edge, so $\min(\hat{w}(A \cup B), \hat{w}(A \cap B)) \geq \xi^2 \hat{w}(A \cup B)$ Finally, the canonical paths have length at most m, and we start the walk from the empty set, which has maximum weight among states of the Markov chain and hence $\hat{\pi}(\emptyset) \geq 2^{-m}$. Plugging into Lemma 1.3, we obtain

$$T_{mix}(\epsilon, \emptyset) = O(m^3 \xi^{-6} \log \epsilon^{-1}).$$

□

1.2.3 From approximate sampling of $\hat{\pi}$ to approximately estimating Z

We know how to get a good estimate of Z, the partition function of the ferromagnetic Ising model, given a good estimate of $\hat{Z}$: simply multiply by $C(\beta, h)$. What is less obvious is how to estimate $\hat{Z}$ given approximate samples from the subgraph world.

In particular, the subgraphs model is not trivially self-reducible, since once an edge has been set to be in or out, that must be further considered when counting the number of odd degree vertices in the reduced problem.

This can be handled, but we will take a different approach:

Fix η and define $\hat{Z}(\xi) := \hat{Z}(\xi, \eta)$. We will define a sequence of ξ values

$$1 = \xi_0 > \xi_1 > \ldots > \xi_r = \xi$$

and then write

$$\hat{Z}(\xi) = \hat{Z}(\xi_0) \prod_{i=1}^{r} \frac{\hat{Z}(\xi_i)}{\hat{Z}(\xi_{i-1})}.$$

Two easy observations:

1) $\hat{Z}(\xi_i)$ is decreasing sequence.

2) $\hat{Z}(\xi_0) = \hat{Z}(1) = \sum_{\ell=0}^{|E|} \binom{|E|}{\ell} \eta^\ell = (1+\eta)^{|E|}$ and therefore is trivially computable.

This leaves us with the job of figuring out how to estimate $\frac{\hat{Z}(\xi_i)}{\hat{Z}(\xi_{i-1})}$.

We will do this by sampling a subgraph A from the subgraphs model with parameters ξ_{i-1} and η and then evaluating the random variable X on that sample, defined as

$$X := \left(\frac{\xi_i}{\xi_{i-1}}\right)^{|odd(A)|}.$$

Observe that

$$\begin{aligned}
E(X) &= \frac{1}{\hat{Z}(\xi_{i-1})} \sum_{A \subseteq E} \hat{w}_{i-1}(A) \left(\frac{\xi_i}{\xi_{i-1}}\right)^{|\mathrm{odd}(A)|} \\
&= \frac{1}{\hat{Z}(\xi_{i-1})} \sum_{k \text{ even}} \sum_{A||\mathrm{odd}(A)|=k} \eta^{|A|} \xi_{i-1}^k \left(\frac{\xi_i}{\xi_{i-1}}\right)^k \\
&= \frac{1}{\hat{Z}(\xi_{i-1})} \sum_{k \text{ even}} \sum_{A||\mathrm{odd}(A)|=k} \eta^{|A|} \xi_i^k \\
&= \frac{\hat{Z}(\xi_i)}{\hat{Z}(\xi_{i-1})}
\end{aligned}$$

Therefore, if we sample from the distribution $\hat{\pi}_{i-1}$ sufficiently many times, and average the values of X on those samples, we will obtain a good estimate of $E(X)$.

Of course, to make sure the number of samples t we need is not too small, and since $X \leq 1$, we will need the expected value of X not to be too small.

If we take

$$\frac{\xi_i}{\xi_{i-1}} = 1 - \frac{1}{n},$$

then

$$\left(\frac{\xi_i}{\xi_{i-1}}\right)^{|\mathrm{odd}(A)|} = \left(1 - \frac{1}{n}\right)^{|\mathrm{odd}(A)|} \in [1/e, 1]$$

so a polynomial number of samples suffices to get $1 + 1/poly(n)$ error. Thus, taking $\xi_i = (1 - 1/n)^i$, for $i = 1, \ldots, cn \ln n$, we can get down from $\xi_0 = 1$ to any $\xi \geq 1/n$. Moreover, from Theorem 1.13, the running time to obtain each sample for values of ξ in this range is polynomial.

Finally, for the case where $0 \le \xi < 1/n$, we show that

$$\frac{\hat{Z}(0)}{\hat{Z}(1/n)} \ge 1/e^2.$$

which allows us to estimate $\hat{Z}(\xi)$ for any ξ in this range by sampling at $\xi = 1/n$. (To estimate $\hat{Z}(0)$, for example, we sample at $\xi = 1/n$, and evaluate the random variable $X = 0^{|\text{odd}(A)|} = \mathbf{1}\{\text{odd}(A) = \emptyset\}$.) It remains to show:

Claim 1.14.

$$\frac{\hat{Z}(0)}{\hat{Z}(1/n)} \ge 1/e^2.$$

Proof. For this, we turn back to the Ising model and recall that $\xi = \tanh(\beta h)$. Since it holds that

$$x \le \frac{\tanh x}{1 - \tanh^2 x} \quad \forall x \ge 0$$

when $\tanh(\beta h) = \xi = 1/n$, this inequality gives

$$\beta h \le \frac{1/n}{1 - 1/n^2} \le \frac{2}{n}.$$

Now consider any configuration σ of the Ising model. and evaluate $w_{\beta,0}(\sigma)/w_{\beta,h}(\sigma)$, where $w_{\beta,h}$ is the weight of σ in the Ising model with parameters β and h when $\beta h \le 2/n$ and $w_{\beta,0}$ is the weight of σ for the Ising model with the same β but $h = 0$. Then

$$\frac{w_{\beta,0}(\sigma)}{w_{\beta,h}(\sigma)} = e^{-\beta h \sum_{u \in V} \sigma(u)} \ge e^{-\frac{2}{n} \sum_{u \in V} \sigma(u)} \ge \frac{1}{e^2}.$$

It follows that

$$\frac{\hat{Z}(0)}{\hat{Z}(1/n)} = \frac{Z(\beta, h = 0)}{Z(\beta, h : \beta h \le 2/n)} \times \frac{C_{\beta,h}}{C_{\beta,0}} \ge \frac{1}{e^2},$$

since $C_{\beta,h} \ge C_{\beta,0}$.

□

References

[JS93] Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the ising model. *SIAM J. Comput.*, 22(5):1087–1116, 1993. 1-1

[NM53] Gordon F Newell and Elliott W Montroll. On the theory of the ising model of ferromagnetism. *Reviews of Modern Physics*, 25(2):353, 1953. 1-2