These notes have not received the scrutiny of publication. They could be missing important references, etc.

Approximate Counting

Introduction

In this lecture we look at how to do approximate counting under Dobrushin condition and the tightness of Dobrushin condition. Recall last time we have defined the Dobrushin influence matrix $R_{n\times n}$:

$$R_{ij} = \sup_{\mathbf{y}_{\sim j}=\mathbf{z}_{\sim j}} |\pi(x_i = \cdot \mid \mathbf{y}) - \pi(x_i = \cdot \mid \mathbf{z})|$$

where π is a probability measure on $\bigotimes_{i=1}^n \Sigma_i$, and proved the following result due to Hayes and Wu:

Theorem 1. *Denote P as the probability transition matrix of Glauber dynamics for a given probability measure π and R as its influence matrix, then the second largest eigenvalue of P satisfies:*

$$1 - \lambda_2(P) \geq \frac{1 - \|R\|_{\text{op}}}{n}.$$

Therefore under the condition "$\|R\|_{\text{op}} < 1 - \delta$" for some small δ, which is called the *Dobrushin uniqueness condition*, we can sample π efficiently using Glauber dynamics P. Today we are going to show that we can also do approximate counting under Dobrushin condition. This also demonstrates the connection between sampling and counting.

1 Approximate Counting

Consider the Ferromagnetic Ising model on a graph G, whose probability distribution takes the following form:

$$\pi_\beta(\mathbf{x}) = \frac{1}{Z_\beta} \exp(\beta \sum_{i\sim j} x_i x_j)$$

where $\mathbf{x} \in \{\pm 1\}^n$, n is the number of vertices of G and $i \sim j$ means there is an edge between vertex i and vertex j. Our goal is to compute $Z_\beta = \sum_{\mathbf{x}\in\{\pm1\}^n} \exp(\beta \sum_{i\sim j} x_i x_j)$. Recall the fact that unless P = NP, there would be no polynomial time algorithm to compute Z_β exactly. However, under some conditions we can compute Z approximately to any precision in polynomial time.

Let d denote the maximum degree of G, i.e., $d = \max_i \sum_j A_{ij}$, where A is the adjacency matrix of graph G. Suppose Dobrushin condition is satisfied: $\beta < 0.99/d$, which implies $1 - \lambda_2(P) = \Theta(1/n)$. We have learned that under this condition we can sample π_β quickly. Today we will show that we can also compute Z_β approximately under Dobrushin condition. Here "approximate count" means that we can compute $\hat{Z}_\beta$ in polynomial time such that $\hat{Z}_\beta \in [(1-\varepsilon)Z_\beta, (1+\varepsilon)Z_\beta]$ for any $\varepsilon > 0$.

1.1 Naive Approach

A naive approach is motivated by reformulating Z_β as:

$$\begin{aligned} Z_\beta &= \frac{2^n}{2^n} \sum_{\mathbf{x}\in\{\pm1\}^n} \exp(\beta \sum_{i\sim j} x_i x_j) \\ &= 2^n \mathbb{E}_{\mathbf{x}\sim\text{Unif}\{\pm1\}^n} \left[\exp(\beta \sum_{i\sim j} x_i x_j) \right]. \end{aligned}$$

Using sample average $\frac{1}{m}\sum_{a=1}^{m}\exp(\beta\sum_{i\sim j}x_i^{(a)}x_j^{(a)})$ to approximate the expectation, where $\mathbf{x}^{(1)},\cdots,\mathbf{x}^{(m)}$ are i.i.d. samples from Unif $\{\pm 1\}^n$, we can get $\hat{Z}_\beta$:

$$\hat{Z}_\beta(m) := 2^n \frac{1}{m}\sum_{a=1}^{m}\exp(\beta\sum_{i\sim j}x_i^{(a)}x_j^{(a)}).$$

By Law of Large Numbers, we know that:

$$\frac{1}{m}\sum_{a=1}^{m}\exp(\beta\sum_{i\sim j}x_i^{(a)}x_j^{(a)}) \longrightarrow \mathbb{E}\left[\exp(\beta\sum_{i\sim j}x_ix_j)\right]$$

as $m\to\infty$. Therefore $\hat{Z}_\beta(m)\to Z_\beta$. The question is that to obtain an ε precision, how lager m needs to take:

Consider the variance of $\hat{Z}_\beta(m)$:

$$\mathrm{Var}\left[\hat{Z}_\beta(m)\right] = 2^{2n}\frac{1}{m}\mathrm{Var}_{\mathbf{x}\sim\mathrm{Unif}\{\pm 1\}^n}(\exp(\beta\sum_{i\sim j}x_ix_j)),$$

where

$$\begin{aligned}\mathrm{Var}_{\mathbf{x}\sim\mathrm{Unif}\{\pm 1\}^n}\left[\exp(\beta\sum_{i\sim j}x_ix_j)\right] &= \mathbb{E}\left[\exp(2\beta\sum_{i\sim j}x_ix_j)\right] - \left(\mathbb{E}\left[\exp(\beta\sum_{i\sim j}x_ix_j)\right]\right)^2 \\ &= \frac{1}{2^n}Z_{2\beta} - \left(\frac{1}{2^n}Z_\beta\right)^2.\end{aligned} \tag{1}$$

By Chebyshev's inequality,

$$\mathbb{P}\left(\left|\hat{Z}_\beta - Z_\beta\right| \geq \varepsilon Z_\beta\right) \leq \frac{\mathrm{Var}\left[\hat{Z}_\beta(m)\right]}{(\varepsilon Z_\beta)^2}.$$

Therefore in order to obtain an ε precision approximation, we need $\mathrm{Var}\left[\hat{Z}_\beta(m)\right] \approx (\varepsilon Z_\beta)^2$.

Example 1. We use 1D Ising model as an example to illustrate sample complexity required of the above approach. We have computed the explicit formula for Z_β in the first lecture:

$$Z_\beta = \sum_{\mathbf{x}\in\{\pm 1\}^n}\prod_{i=1}^{n-1}\exp(\beta x_ix_{i+1}) = 2^n\cosh(\beta)^{n-1}$$

Compute the ratio of the two terms in (1):

$$\frac{Z_{2\beta}/2^n}{(Z_\beta/2^n)^2} = \frac{2^nZ_{2\beta}}{Z_\beta^2} = \frac{\cosh(2\beta)^{n-1}}{cosh(\beta)^{2n-2}} = \left(\frac{\cosh^2\beta+\sinh^2\beta}{\cosh^2\beta}\right)^{n-1} = \left(1+\tanh^2\beta\right)^{n-1}.$$

Therefore,

$$\begin{aligned}\mathrm{Var}\left[\hat{Z}_\beta(m)\right] &= \frac{2^{2n}}{m}\left[1+\left(1+\tanh^2\beta\right)^{n-1}\right]\frac{Z_\beta^2}{2^{2n}} \\ &= \frac{1+\left(1+\tanh^2\beta\right)^{n-1}}{m}Z_\beta^2.\end{aligned}$$

In order to get a $(1\pm\varepsilon)$ approximation of Z_β, i.e. $\mathrm{Var}\left[\hat{Z}_\beta(m)\right] \approx (\varepsilon Z_\beta)^2$, m needs to be of order $\frac{1+\left(1+\tanh^2\beta\right)^{n-1}}{\varepsilon^2}$, which depends on n exponentially, hence is a very slow method.

1.2 Simulated Annealing

Can we get a faster algorithm to compute Z_β? Take another look at the above method, what we actually do is to write $Z_\beta = \frac{Z_\beta}{Z_0} Z_0\ Z_0$, where $Z_0 := \sum_{\mathbf{x}\in\{\pm 1\}^n} 1 = 2^n$, and compute the ratio Z_β/Z_0 by sampling from Unif $\{\pm 1\}^n$. This method is not efficient since Z_0 is far from Z_β. We can get a more efficient method by varying the temperature β (actually the inverse of temperature) slowly. To be more specific, we can write Z_β as the following:

$$Z_\beta = \frac{Z_\beta}{Z_{\beta\frac{k-1}{k}}} \cdot \frac{Z_{\beta\frac{k-1}{k}}}{Z_{\beta\frac{k-2}{k-1}}} \cdots \frac{Z_{\beta\frac{1}{k}}}{Z_0} Z_0. \tag{2}$$

Define $\delta = \beta/k$. We need to estimate $Z_\beta/Z_{\beta-\delta}$, then we can estimate $Z_{\beta-(l-1)\delta}/Z_{\beta-l\delta}$ for any $2 \le l \le k$ similarily. Using the same idea as in the above method, we can reformulate $Z_\beta/Z_{\beta-\delta}$ as the following:

$$\begin{aligned}
\frac{Z_\beta}{Z_{\beta-\delta}} &= \frac{1}{Z_{\beta-\delta}} \sum_{\mathbf{x}\in\{\pm 1\}^n} \exp(\beta \sum_{i\sim j} x_i x_j) \\
&= \frac{1}{Z_{\beta-\delta}} \sum_{\mathbf{x}\in\{\pm 1\}^n} \exp((\beta-\delta) \sum_{i\sim j} x_i x_j) \cdot \exp(\delta \sum_{i\sim j} x_i x_j) \\
&= \mathbb{E}_{\mathbf{x}\sim\pi_{\beta-\delta}} \left[\exp(\delta \sum_{i\sim j} x_i x_j) \right]
\end{aligned}$$

Define $\hat{y}_{\beta-\delta}(m) = \frac{1}{m}\sum_{a=1}^m \exp(\delta \sum_{i\sim j} x_i^{(a)} x_j^{(a)})$ where $\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(m)}$ are i.i.d. samples from $\pi_{\beta-\delta}$. Note that if π_β satisfies Dobrushin cnodition, i.e. $\beta/d < 0.99$, then it is clear that $\pi_{\beta-\delta}$ also satisfies Dobrushin condition, thus we can get samples $\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(m)}$ from $\pi_{\beta-\delta}$ quickly.

Choose $\delta < 1/nd$, so that $\delta \sum_{i\sim j} x_i^{(a)} x_j^{(a)} \in [-1, 1]$. Then $\exp(\delta \sum_{i\sim j} x_i x_j)$ is a bounded random variable and its variance is then bounded by some constant C(see Popoviciu's inequality). Therefore

$$\begin{aligned}
\operatorname{Var}\left[\hat{y}_{\beta-\delta}(m)\right] &= \frac{1}{m} \operatorname{Var}_{\mathbf{x}\sim\pi_{\beta-\delta}} \left[\exp(\delta \sum_{i\sim j} x_i^{(a)} x_j^{(a)}) \right] \\
&\le \frac{1}{m} \cdot \mathrm{C}\,.
\end{aligned}$$

In order to get a $(1 \pm \varepsilon)$ approximation of $Z_\beta/Z_{\beta-\delta}$, m needs to be $\Theta(\frac{1}{\varepsilon^2})$.

Similarily we can compute other terms in (2) approximately:

$$\hat{y}_{\beta-l\delta} \in \left[(1-\varepsilon)\frac{Z_{\beta-(l-1)\delta}}{Z_{\beta-l\delta}}, (1+\varepsilon)\frac{Z_{\beta-(l-1)\delta}}{Z_{\beta-l\delta}} \right], \qquad \forall\, 2 \le l \le k.$$

Finally we get the approximation $\hat{Z}_\beta$:

$$\hat{Z}_\beta = \hat{y}_{\beta-\delta} \cdot \hat{y}_{\beta-2\delta} \cdots \hat{y}_\delta \cdot 2^n.$$

We further write it as

$$\log \hat{Z}_\beta = \log \hat{y}_{\beta-\delta} + \cdots + \log \hat{y}_\delta + n \log 2.$$

According to the property of $\hat{y}_{\beta-l\delta}$,

$$\begin{aligned}
\log \hat{Z}_\beta &= \left(\log \frac{Z_\beta}{Z_{\beta-\delta}} \pm \varepsilon \right) + \left(\log \frac{Z_{\beta-\delta}}{Z_{\beta-2\delta}} \pm \varepsilon \right) + \cdots + n \log 2 \\
&= \log Z_\beta \pm k\varepsilon.
\end{aligned}$$

Denote $\varepsilon' = k\varepsilon$, we can get a $(1 \pm \varepsilon')$ approximation of $\hat{Z}_\beta$ with total sample size $\Theta(k\varepsilon^{-2})$. Since $k = \beta/\delta$ and we choose δ to be less than $\frac{1}{nd}$, the total sample size depends on n polynomially.

Remark 1. There are some different approches to approximately compute the partition function. For example we can write Z_β as the following:

$$\begin{aligned} Z_\beta &= \sum_{\mathbf{x}\in\{\pm 1\}^n} \exp(\beta \sum_{i\sim j} x_i x_j) \\ &= \sum_{x_1\in\{\pm 1\}} \sum_{\mathbf{x}_{\sim 1}} \exp(\beta \sum_{i\sim j} x_i x_j) \\ &= Z_\beta(x_1=1) + Z_\beta(x_1=-1) \\ &= Z_\beta(x_1=1)\cdot \frac{Z_\beta(x_1=1)+Z_\beta(x_1=-1)}{Z_\beta(x_1=1)}. \end{aligned}$$

Define $\pi_\beta(x_1=1) = \frac{Z_\beta(x_1=1)}{Z_\beta(x_1=1)+Z_\beta(x_1=-1)}$, then we can compute Z_β by computing $Z_\beta(x_1=1)$ and $\pi_\beta(x_1=1)$. Continue decomposing $Z_\beta(x_1=1)$ in terms of x_2 in the same way, we can get the approximation of Z_β.

2 Tightness of Dobrushin

We show the tightness of Dobrushin condition in Curie-Weiss model, where every spin interacts with every other spin with the same strength, i.e. Ising model on a complete graph. The distribution takes the following form:

$$\pi_\beta(\mathbf{x}) = \frac{1}{Z_\beta} \exp\left(\frac{\beta}{2n}\left(\sum_{i=1}^n x_i\right)^2\right).$$

If Dobrushin condition $\beta < 0.99$ is satisfied, then we know that Gibbs sampler mixes quickly. However if β is larger than 1, say $\beta = 1.0001$, we would get stuck by "torpid mixing", and the eigengap in this scenario is about $1-\lambda_2 = \exp(-\Theta(n))$. Hence Dobrushin condition is tight in the sense of getting samples efficiently by Glauber dynamics.

This phenomenon is related to the property that the model undergoes a phase transition when β equals 1. Define $M = \sum_{i=1}^n x_i$. Note that we have:

$$Z_\beta = \sum_{m=-n}^{n} \exp\left(\frac{\beta}{2n} m^2\right) \cdot \#\left\{\mathbf{x}\in\{\pm 1\}^n \mid M = m\right\}.$$

Since $\#\left\{\mathbf{x}\in\{\pm 1\}^n \mid M = m\right\} = \left\{ {n \atop (n+m)/2} \right\}$, we can compute

$$\begin{aligned} \mathbb{P}\left(M=m\right) &\propto \exp\left(\frac{\beta}{2n}m^2 + \log \binom{n}{(n+m)/2}\right) \\ &= \exp(n f_n(\tilde{m})), \end{aligned}$$

then the where $\tilde{m} = m/n$ is the average magnetization and $f_n(\tilde{m}) = \frac{\beta}{2}\tilde{m}^2 + \frac{1}{n}\log\binom{n}{n(1+\tilde{m})/2}$. we are intereste in calculating the average magnetization $m^\star = \arg\max \mathbb{P}(\tilde{m})$ as $n\to\infty$. By stirling's formula, we have a closed form of $f_n(\tilde{m})$, We can see there is a phase transition in $\beta = 1$. In the case of high temperature ($\beta < 1$), the only maximum of $\mathbb{P}(\tilde{m})$ is zero and when temperature is low ($\beta > 1$) there will be two maxima symmetrically distributed on both sides of 0. We will discuss this phase transition process in detail in the next lecture.