Randomness and Computation 2023
Tutorial 8 (week 10)
Solution

1. **Gibbs sampling for the Ferromagnetic Ising model.** A spin system is a probability distribution on $\mathcal{X} = \{-1, +1\}^V$, where V vertices of graph $G = (V, E)$. The value $\gamma(v)$ is called the spin of v. The Ising model is used by physicist as a classical approximation to model magnetic properties of materials. The interpretation is that magnets have up or down orientation, encoded by the $+1$ and -1, and are placed in the vertices of the graph, where the edges will encode the interaction between magnets. The nearest-neighbor ferromagnetic Ising model is one of the most widely studied spin system. The energy of a configuration γ is defined to be

$$H(\gamma) = -\sum_{\substack{v,w \in V \\ v \sim w}} \sigma(v)\sigma(w), \tag{1}$$

where the energy increases with the number of pairs of neighbors whose spin disagree. The Gibbs distribution corresponding to energy H is the probability distribution μ on $\mathcal{X}$ defined by

$$\mu(\sigma) = \frac{1}{Z(\beta)} e^{-\beta H(\sigma)},$$

where $\beta \geq 0$ is a parameter related to the inverse temperature and $Z(\beta) = \sum_{\sigma \in \mathcal{X}} e^{-\beta H(\sigma)}$ is a normalization term called partition function, that plays an important role in the physical description of the system.

   (a) Show that for $\beta = 0$ the distribution μ is nothing else than the uniform distribution over all spin configurations.

   (b) Show that for $\beta = \infty$, μ is the uniform distribution over the set of configuration σ that minimize $H(\sigma)$, i.e., with probability $1/2$ all spins aligned to $+1$ and with probability $1/2$ all spins aligned to -1.

   (c) Show that the transition probability of the Glauber dynamics selecting uniformly at random a vertex w from V reads:

   $$P(\sigma_t, \sigma_{t+1}^{\pm 1_w}) = (1 + \sigma_{t+1}(w) \tanh(\beta S(\sigma_t, w))/2, \tag{2}$$

   where $S(\sigma, w) = \sum_{u: u \sim w} \sigma(u)$, σ_t is the initial spin configuration and $\sigma_{t+1}^{\pm 1_w}$ is the updated configuration with $\sigma_{t+1}^{\pm 1_w}(x) = \sigma_t(x) \forall x \neq w$ and $\sigma_{t+1}^{+1_w}(w) = +1$ or $\sigma_{t+1}^{-1_w}(w) = -1$.

   (d) Prove that the corresponding Markov chain is irreducible and aperiodic.

   (e) Show it satisfies detailed balance.

Solution

   (a) It is easy to see that for $\beta = 0$ we have $e^{-\beta H(\sigma)} = 1$ and $Z(\beta) = |\mathcal{X}| = 2^n$, leading to the uniform distribution over the spin configurations $\mu(\sigma) = 1/2^n$.

(b) It is easy to see that the two solutions σ_+ of all spins $+1$ and σ_- of all spins -1 minimize the energy $H(\sigma_\pm) = -|E|$. One can then re-write $\mu(\sigma)$ as:

$$\mu(\sigma) = \frac{e^{-\beta H(\sigma)}}{2e^{\beta|E|} + \sum_{\sigma\in\mathcal{X}\setminus\{\sigma_\pm\}} e^{-\beta H(\sigma)}} = \frac{e^{-\beta(H(\sigma)+|E|)}}{2 + \sum_{\sigma\in\mathcal{X}\setminus\{\sigma_\pm\}} e^{-\beta(H(\sigma)+|E|)}} \tag{3}$$

which for $\beta = \infty$ has value $1/2$ for $\sigma_\pm$ (as $H(\sigma_\pm) + |E| = 0$) and 0 for the other spin configurations (as $H(\sigma)+|E| > 0$ for $\sigma \neq \sigma_\pm$). This case correspond the zero temperature Gibbs distribution that is the mixture of the states that minimize the energy of the system.

(c) Taking into consideration that the two update spin configurations only differ on vertex w we can write:

$$H(\sigma) = H(\sigma)_{V\setminus w} - \sigma(w) \sum_{x\in V:x\sim w} \sigma(u) = H_{V\setminus w} - \sigma(w)S(\sigma, w). \tag{4}$$

where $H_{V\setminus w}$ is the term of the energy that does has no dependency on the vertex w. we can then write,

$$P(\sigma_t, \sigma_{t+1}^{\pm 1}) = \frac{e^{-\beta H_{V\setminus v}} e^{\beta\sigma(v)S(\sigma,v)}}{e^{-\beta H_{V\setminus v}} e^{\beta S(\sigma,v)} + e^{-\beta H_{V\setminus v}} e^{-\beta S(\sigma,v)}}, \tag{5}$$

which can be simplified to

$$P(\sigma_t, \sigma_{t+1}^{\pm 1_w}) = \frac{e^{\beta\sigma(v)S(\sigma,v)}}{e^{\beta S(\sigma,v)} + e^{-\beta S(\sigma,v)}}, \tag{6}$$

and using the relation

$$1 \pm \tanh x = 1 \pm \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{2e^{\pm x}}{e^x + e^{-x}} \tag{7}$$

leads to

$$P(\sigma_t, \sigma_{t+1}^{\pm 1_w}) = (1 + \sigma(v)\tanh(\beta S(\sigma, v))/2. \tag{8}$$

(d) The argument for irreversibility is similar to the ones done before. Basically any spin configuration can reach any other via a one spin change at a time, where each transition has at least a non-zero probability for finite β and therefore the full path connecting two configurations has always a non-zero probability. Aperiodicity is obtained by having a non-zero probability of remaining in the initial state.

(e) Taking into consideration that spin configuration x and y only differ on vertex w we can write:

$$\pi(x) = \frac{1}{Z(\beta)} e^{-\beta H_{V\setminus w}} e^{\sigma_x(w)S(\sigma_x,w)}, \tag{9}$$

and

$$\pi(y) = \frac{1}{Z(\beta)} e^{-\beta H_{V\setminus w}} e^{\sigma_y(w)S(\sigma_y,w)}, \tag{10}$$

where we have the relation $S(\sigma_x, w) = S(\sigma_y, w)$ as the configuration are equal outside w. We can therefore write:

$$\pi(x)P_{x,y} = \frac{1}{Z}e^{H_{V\setminus w}}e^{\beta\sigma_x(w)S(\sigma_x,w)}\frac{e^{\beta\sigma_y(w)S(\sigma_y,w)}}{e^{\beta S(\sigma_y,w)} + e^{-\beta S(\sigma_y,w)}} \tag{11}$$

where using $S(\sigma_x, w) = S(\sigma_y, w)$, leads to

$$\pi(x)P_{x,y} = \frac{1}{Z}e^{H_{V\setminus w}}e^{\beta\sigma_y(w)S(\sigma_y,w)}\frac{e^{\beta\sigma_x(w)S(\sigma_x,w)}}{e^{\beta S(\sigma_x,w)} + e^{-\beta S(\sigma_x,w)}} = \pi(y)Py, x \tag{12}$$

2. **Path coupling for the Ferromagnetic Ising model**.

   Let's define the distance ρ on $\mathcal{X}$ by

   $$\rho(\sigma, \tau) = \frac{1}{2}\sum_{u\in V}|\sigma(u) - \tau(u)|,$$

   which is easy to see that quantifies how many spins are different.

   We are going to follow the standard technique for proving rapid mixing presented in lecture 19 of the course, where we design a coupling consisting on applying the same update to both chains X_t and Y_t. Let's consider σ and τ such that $\rho(\sigma, \tau) = 1$, i.e., they agree everywhere except in vertex v. Let's define $N(v) = \{u : u \sim v\}$ to be the set of neighbors vertices of v.

   (a) Explain why when the vertex w selected during the update satisfies $w = v$ we have $\rho(X_{t+1}, Y_{t+1}) = 0$.

   (b) Justify why when $w \notin N(v) \cup \{v\}$, then the distance does not change, i.e., $\rho(X_{t+1}, Y_{t+1}) = 1$.

   (c) Show that the probability of $\rho(X_{t+1}, Y_{t+1}) = 2$, when $w \in N(v)$, satisfies the condition

   $$P(\rho = 2|w \in N(v)) = |p(\tau, w) - p(\sigma, w)|.$$

   (d) Using the relation $\tanh(\beta(x+1)) - \tanh(\beta(x-1)) \leq \tanh\beta$ prove the relation

   $$E[\rho_{t+1}|\rho_t = 1] \leq 1 - \frac{1 - \Delta\tanh(\beta)}{n} = 1 - \frac{c(\beta)}{n} \leq e^{-c(\beta)/n}.$$

   (e) Discuss under which conditions of Δ and $\tanh(\beta)$ this Glauber dynamics for the Ferromagnetic Ising model shows rapid mixing.

   Solution:

   (a) Because chains X_t and Y_y respective configurations γ and τ only differ on vertex v, when the update vertex $w = v$ the update follow by both chains is exactly the same. Therefore, we replace the configuration of the vertex v, which was were X_t was different Y_t, by the same value on both chains. Therefore we decrease the distance between the chains to zero, i.e., $\rho(X_{t+1}, Y_{t+1}) = 0$.

(b) When $w \notin N(v) \cup \{v\}$, w and all its neighbor vertices have the same value on both chains X_t and Y_t. Therefore, the coupled Glauber dynamics will generate the same update for w on both chains. Because $\sigma(w) = \tau(w)$ already, the distance neither decrease or increase, i.e., $\rho(X_{t+1}, Y_{t+1}) = 1$.

(c) The scenario $w \in N(v)$ has two possible scenarios, if the update of w is the same value the distance is preserved ($\rho(X_{t+1}, Y_{t+1}) = 1$), but with some non-zero probability the update of w can be different for both chains, with increases by one vertex the distance, leading to $\rho(X_{t+1}, Y_{t+1}) = 2$. By inspection of the update rule above (2), and using $U \in [0, 1]$ as the random number used to decide the update of both chains using rule (2), it is not hard to see that:

- Scenario I: corresponds to either $U \leq \min\{P(\sigma_t, \sigma_{t+1}^{+1_w}), P(\tau_t, \tau_{t+1}^{+1_w})\}$ or $\max\{P(\sigma_t, \sigma_{t+1}^{+1_w}), P(\tau_t, \tau_{t+1}^{+1_w})\} \leq U$, where the update reads the same for both chains.
- Scenario II: corresponds to $\min\{P(\sigma_t, \sigma_{t+1}^{+1_w}), P(\tau_t, \tau_{t+1}^{+1_w})\} \leq U \leq \max\{P(\sigma_t, \sigma_{t+1}^{+1_w}), P(\tau_t, \tau_{t+1}^{+1_w})\}$.

Remark that the probability of Scenario II is

$$\begin{aligned} P(\rho = 2 | w \in N(v)) &= \max\{P(\sigma_t, \sigma_{t+1}^{+1_w}), P(\tau_t, \tau_{t+1}^{+1_w})\} - \min\{P(\sigma_t, \sigma_{t+1}^{+1_w}), P(\tau_t, \tau_{t+1}^{+1_w})\} \\ &= |P(\sigma_t, \sigma_{t+1}^{+1_w}) - P(\tau_t, \tau_{t+1}^{+1_w})|. \end{aligned}$$

(d) Our analysis for the coupling of two chains with distance $\rho(X_t, Y_t) = 1$, leads to:

$$E[\rho_{t+1} | \rho_t = 1] = 1 - \frac{1}{n} - \frac{2}{n} \sum_{w \in N(v)} |P(\sigma_t, \sigma_{t+1}^{+1_w}) - P(\tau_t, \tau_{t+1}^{+1_w})|,$$

where the second term on the left side of equality correspond to the probability of $w = v$ and the last term is the probability that we select a neighbor of v and that we end in case II described above. Because when $w \in N(v)$, all the neighbors of w have the same spin except for v where the values are opposite, we have

$$|P(\sigma_t, \sigma_{t+1}^{+1_w}) - P(\tau_t, \tau_{t+1}^{+1_w})| = \frac{1}{2} \left(\tanh(\beta(x+1)) - \tanh(\beta(x-1))\right),$$

where $x = \sum_{u: u \sim w \setminus \{v\}} \sigma(u)$, is the sum of the spins of all neighbors of w except v. Using the relation $\tanh(\beta(x+1)) - \tanh(\beta(x-1)) \leq \tanh\beta$, and the fact that any vertex has at most Δ neighbors, one can further simplify to:

$$E[\rho_{t+1} | \rho_t = 1] = 1 - \frac{1 - \Delta \tanh(\beta)}{n} = 1 - \frac{c(\beta)}{n} \leq e^{-c(\beta)/n}.$$

(e) The previous result is equivalent to stating that for every two chains Z_i and Z_{i+1} having a distance of 1, we have $\mathbf{E}[d(Z'_{i+1}, Z'_i)] \leq e^{-c(\beta)/n}$. Following the discussion on Lecture 19, we can construct a path of of d_t chains Z_i between X_t and Y_t, which allows us to write

$$\mathbf{E}[d_{t+1} | d_t] = \sum_{i=1}^{d_t} \mathbf{E}[d(Z'_{i+1}, Z'_i)] \leq e^{-c(\beta)/n} d_t,$$

which leads to $\mathbf{E}[d_{t+1}] = \mathbf{E}[\mathbf{E}[d_{t+1}|d_t]] =\leq e^{-c(\beta)/n}\mathbf{E}[d_t]$ and by iteration to $\mathbf{E}[d_T] \leq e^{-c(\beta)T/n}d_0 \leq e^{-c(\beta)T/n}n$, where the condition $c(\beta) > 0$ is necessary to prove contraction of the distance with t. This leads to

$$\tau(\epsilon) = \frac{n}{c(\beta)}\left(\log n + \log(1/\epsilon)\right),$$

where the condition $c(\beta) > 0$ is necessary to reach rapid mixing, which corresponds to $\Delta \tanh \beta < 1$.

3. **Path coupling for Gibbs distribution on independent sets.** Let $G = (V, E)$ be a graph with maximum degree Δ, Ω be the set of independent sets on G, and $x \in \{0,1\}^V$ a binary encoding of the vertices composing a given independent set. Let $\pi(x)$ the Gibbs distribution on independent sets:

$$\pi(x) = \begin{cases} \frac{\lambda^{|x|}}{Z(\lambda)} & \text{if } x(v)x(w) = 0 \quad \forall \{v,w\} \in E \\ 0 & \text{Otherwise.} \end{cases} \tag{13}$$

where $|x|$ is the Hamming weight of configuration x, i.e., $|x| = \sum_{v \in V} x(v)$ and $Z(\lambda) = \sum_{x \in X} \lambda^{|x|}$ normalizes π. The Glauber dynamics updates configuration X_t to a new configuration X_{t+1} by first selecting a vertex $v \in V$ uniformly at random and then implementing an update. Using the general definition of a Glauber dynamic update, show that we obtain the following update rule:

First, set $X_{t+1}(w) = X_t(w) \quad \forall w \neq v$. Then, if exist $w' \in N(v)$, where $N(v)$ is the neighborhood of v, such that $X_t(w') = 1$.

- we set (M1) [1] $X_{t+1}(v) = 0$
- otherwise (M2)

$$X_{t+1}(v) = \begin{cases} 1 & \text{with probability } \lambda/(1+\lambda), \\ 0 & \text{with probability } 1/(1+\lambda). \end{cases} \tag{14}$$

(a) Let's define two chains X_t and Y_t that we will couple via the application of the exact same update rule at each step. Let's define the distance $\rho(X, Y)$ between two independent sets by the amount of vertices that are different, i.e, $\rho(X, Y) = \sum_{i \in V} |x(i) - y(i)|$. Assume that X_t and Y_t differ in a single vertex v:

i. Explain why when the vertex w selected during the update satisfies $w = v$ we have $\rho(X_{t+1}, Y_{t+1}) = 0$.
Solution: If we select the vertex v where X_t and Y_t differ, all neighbors are the same in both chains, which implies the chain applies same move to both. Because either $X_t(v) = 1(Y_t(v) = 0)$ or $Y_t(v) = 1(X_t(v) = 10$, all its neighbors have to be 0. Therefore the MC will chose move M2. Whatever is the update, now we have $X_t(v) = Y_t(v)$ and $\rho(X_{t+1}, Y_{t+1}) = 0$.

[1] Remark that the labeling (M1) and (M2) is used to facilitate the discussion of the solution.

ii. Justify why when $w \notin N(v) \cup \{v\}$, then the distance does not change, i.e., $\rho(X_{t+1}, Y_{t+1}) = 1$.

Solution: All neighbors of $w \notin N(v) \cup \{v\}$ are equal on both chains, which lead to the MC to implement the same update (M1 or M2 with same output) on both chains. Because initially $X_t(w) = Y_t(w)$ and the outcome also satisfies $X_{t+1}(w) = Y_{t+1}(w)$ we have that the distance does not change, i.e., v remains the only vertex where they differ.

iii. Show that the probability of $\rho(X_{t+1}, Y_{t+1}) = 2$, when $w \in N(v)$, is always no larger than $\lambda/(1+\lambda)$.

Solution: If the selected vertex w is a neighbor of v. Assume without lost of generality that $X_t(v) = 1 (Y_t(v) = 0)$. What may happen is that the MC implements move $M1$ to X_t and M_2 to Y_t potentially leading to $Y_{t+1}(w) = 1 \neq X_{t+1}(w)$. Because the two chain were differing on v already, now $\rho(X_{t+1}, Y_{t+1}) = 2$. This can only happen if all neighbors of w in Y_t are 0 and the move M2 select output 1 for w, which has total probability $\leq \lambda/(1+\lambda)$. An equivalent argument works for the case $X_t(v) = 0 (Y_t(v) = 1)$, exchanging the roles of X and Y.

(b) For which condition between the parameter λ and maximum degree of the graph Δ we have $E(\rho(X_{t+1}, Y_{t+1})) \leq 1 - c(\lambda)/n \leq e^{-c(\lambda)/n}$ with $c(\lambda) > 0$ when $\rho(X_t, Y_t) = 1$?

Solution From the derivation above we can show that

$$\rho(X_{t+1}, Y_{t+1}) \leq 1 - \frac{1}{n} + \frac{\Delta}{n}\frac{\lambda}{1+\lambda},$$

where the negative term corresponds to the probability of selecting vertex v where both are different and the last term correspond to the case of selecting a neighbor of v (probability $\leq \frac{\Delta}{n}$) followed by a potential update to 1 on one chain and not the other. This can be simplified to

$$\rho(X_{t+1}, Y_{t+1}) \leq 1 - \frac{1}{n}\frac{1-\lambda(\Delta-1)}{1+\lambda}.$$

If $\lambda < (\Delta - 1)^{-1}$ then we have

$$E[\rho(X_{t+1}, Y_{t+1})] \leq 1 - \frac{c(\lambda)}{n} \leq e^{-c(\lambda)/n},$$

with $c(\lambda) = 1 - \lambda(\Delta - 1)$.

(c) Prove that the mixing time will satisfy $t_{\text{mix}}(\epsilon) \leq \frac{n}{c(\lambda)}(\log n + \log \epsilon^{-1})$.

Solution If we have now $\rho(X_{t+1}, Y_{t+1}) = d_t$ we can always construct a path of d_t chains Z_i, where $Z_1 = X_t$ and $Z_{d_t} = Y_t$, such that $\rho(Z_i, Z_{i+1}) = 1$. Coupling all chain to the same MC we obtain

$$E[d_{t+1}|d_t] = E[\rho(X_{t+1}, Y_{t+1})] = E[\sum_{i=1}^{d_t} \rho(Z_i, Z_{i+1})] = \sum_{i=1}^{d_t} E[\rho(Z_i, Z_{i+1})] \leq d_t e^{-c(\lambda)/n}.$$

Using the conditional expectation equality we can then obtain

$$E[d_{t+1}|d_t] = E[E[d_{t+1}|d_t]] = E[d_t]e^{-c(\lambda)/n},$$

which by induction leads to

$$E[d_{t+1}] \leq d_0 e^{-c(\lambda)t/n} \leq n e^{-c(\lambda)t/n}.$$

The coupling lemma states that if we can prove $\Pr(X_T \neq Y_T | X_0 = x, Y_0 = y) \leq \epsilon$ then the mixing time satisfies $\tau(\epsilon) \leq T$.
Because $X_t = Y_t$ if and only if $d_t = 0$, we have

$$\Pr(X_T \neq Y_T | X_0 = x, Y_0 = y) = \Pr(d_t \geq 1 | X_0 = x, Y_0 = y) \leq E[d_t | X_0 = x, Y_0 = y] \leq n e^{-c(\lambda)t/n}.$$

By choosing

$$T = \frac{n}{c(\lambda)}(\log(n/\epsilon))$$

we ensure $\Pr(X_T \neq Y_T) \leq \epsilon$ and therefore the mixing of the chain.

Raul Garcia-Patron Sanchez