Yizhong Hu Summer 2023 Running Document

Yizhong Hu

June 11, 2023

1 Jun 1 - Jun 5, 2023

1.1 Model Description

We consider the Hamiltonian $\mathcal{H} : \mathcal{X}^{\kappa+1} \times \mathcal{Y}^{\kappa} \mapsto \mathbb{R}$

$$\mathcal{H}(x_0, x_1, \ldots, x_\kappa, y_{01}, \ldots, y_{0\kappa}) = \frac{\beta}{2} \sum_{i=1}^{\kappa} x_0 x_i y_{0i} + B x_0$$

where x_0 is spin state of the root node, x_i is the spin state of the i-th leaf node, y_{0i} is the interaction between 0 and i. All of them take values in $\mathcal{X} = \mathcal{Y} = \{-1, 1\}$.

When $\kappa = 2$,

$$\mathcal{H}(x_0, x_1, x_2, y_{01}, y_{02}) = \frac{\beta}{2}(x_0 x_1 y_{01} + x_0 x_2 y_{02}) + B x_0$$

We define the state s in state space $\mathbb{S} = \mathcal{X}^3 \times \mathcal{Y}^2$, and the random variable S a random variable on $\mathbb{S}$. Given η the i.i.d Bernoulli(1/2) distribution on $\mathbb{S}$, we want to investigate the distribution $\mu(\cdot)$ that maximizes

$$\max_{\mu(\cdot)} \{\mathbb{E}_\mu[\mathcal{H}(S)] - [H(\mu \| \eta) + H(\mu_{01} \| \eta_{01})]\}$$

1.2 Analytical analysis

First, we need to find the distribution $\mu : \mathcal{X}^3 \times \mathcal{Y}^2 \mapsto [0, 1]$. Since the input is discrete, μ can be rewritten as a vector on $[0, 1]^{2^5}$. We will denote each component as $\mu(s)$, with $s \in \mathbb{S}$

The marginal distribution is therefore

$$\mu_{01}(y_{01}) = \sum_{x_1, x_2, x_3, y_{02} \in \{-1, 1\}} \mu(s)$$

Rewriting each term,

$$\mathbb{E}_\mu[\mathcal{H}(S)] = \sum_{s\in\mathbb{S}} \mu(s)\mathcal{H}(s)$$
$$H(\mu\|\eta) = \sum_{s\in\mathbb{S}} \mu(s)\log\frac{\mu(s)}{\eta(s)}$$
$$H(\mu_{01}\|\eta_{01}) = \sum_{y_{01}\in\mathcal{Y}} \mu_{01}(y_{01})\log\frac{\mu_{01}(y_{01})}{\eta_{01}(y_{01})}$$

Note that since η is uniform, the relative entropies can be written directly in terms of their entropies:

$$H(\mu\|\eta) = H(\mu) - \log|\mathbb{S}|$$
$$H(\mu_{01}\|\eta_{01}) = H(\mu_{01}) - \log 2$$

Rewriting the target,

$$\max_{\mu(\cdot)}\left[\sum_{s\in\mathbb{S}}\mu(s)\mathcal{H}(s) - \sum_{s\in\mathbb{S}}\mu(s)\log\mu(s) + \sum_{y_{01}\in\mathcal{Y}}\mu_{01}(y_{01})\log\mu_{01}(y_{01})\right]$$

The corresponding Lagrange multiplier (constrained on μ being normalized) is

$$\mathcal{L} = \left[\sum_{s\in\mathbb{S}}\mu(s)\mathcal{H}(s) - \sum_{s\in\mathbb{S}}\mu(s)\log\mu(s) + \sum_{y_{01}\in\mathcal{Y}}\mu_{01}(y_{01})\log\mu_{01}(y_{01})\right] + \lambda\left[\sum_{s\in\mathbb{S}}\mu(s) - 1\right]$$

taking the gradients gives

$$\frac{\partial\mathcal{L}}{\partial\mu(s)} = \mathcal{H}(s) - \log\mu(s) + \log\mu_{01}(y_{01}) + \lambda = 0 \tag{1}$$
$$\frac{\partial\mathcal{L}}{\partial\lambda} = \sum_{s\in\mathbb{S}}\mu(s) - 1 = 0 \tag{2}$$

where y_{01} in Eq.(1) represents the value of y_{01} in s.

Eq.(1) gives the form of μ in terms of a conditional distribution, as seen below:

$$\log\mu(s) - \log\mu_{01}(y_{01}) = \mathcal{H}(s) + \lambda$$
$$\log\frac{\mu(s)}{\mu_{01}(y_{01})} = \mathcal{H}(s) + \lambda$$
$$\frac{\mu(s)}{\mu_{01}(y_{01})} = e^\lambda e^{\mathcal{H}(s)}$$

The left-hand side is a conditional distribution:

$$\mathbb{P}_\mu(S = s|Y_{01} = y_{01}) = \frac{1}{Z}\exp\left[\frac{\beta}{2}(x_0x_1y_{01} + x_0x_2y_{02}) + Bx_0\right] \tag{3}$$

for some $Z = e^{-\lambda}$ = normalizing constant.

Since $\frac{\partial \mathcal{L}}{\partial \mu(s)} = 0$ for all μ_{01} values, and all such optimum points form a connected curve, we think that it is reasonable to hypothesize that any choice of μ_{01} satisfies optimality. Since y_{01} can only take two values, the distribution can be characterized by a single real value α:

$$\mathbb{P}_\mu(Y_{01} = y_{01}) = \frac{e^{\alpha y_{01}}}{e^{\alpha} + e^{-\alpha}},$$

and the entire distribution becomes

$$\mu(s) = \frac{1}{Z} \exp\left[\frac{\beta}{2}(x_0 x_1 y_{01} + x_0 x_2 y_{02}) + B x_0 + \alpha y_{01}\right]$$

1.3 Numerical Analysis

On the other hand, this problem can be solved numerically, formulated as a constrained optimization:

$$\begin{aligned} \text{maximize} \quad & \sum_{s \in \mathbb{S}} \mu(s)\mathcal{H}(s) - \sum_{s \in \mathbb{S}} \mu(s) \log \mu(s) + \sum_{y_{01} \in \mathcal{Y}} \mu_{01}(y_{01}) \log \mu_{01}(y_{01}) \\ \text{subject to} \quad & \sum_{s \in \mathbb{S}} \mu(s) = 1 \end{aligned}$$

Since entropy calculations really don't like getting negative values, we will represent μ as an exponential:

$$\mu(s) = e^{x(s)}$$

Note that since $\exp : \mathbb{R} \mapsto (0, \infty)$ is bijective, we do not lose any generality.

A proximal point optimization method will be used for the optimization. For the following optimization problem

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{g}(\mathbf{x}) = \mathbf{0}, \end{aligned}$$

we have the following iterative process to obtain an optimum:

$$\begin{aligned} \mathbf{x}^{(t+1)} &= \text{argmax}_{\text{x}} f(\text{x}) + \lambda^{(t)} \cdot \mathbf{g}(\mathbf{x}) - \frac{1}{2}(\mathbf{g}(\mathbf{x}))^2 \\ \lambda^{(t+1)} &= \lambda^{(t)} - \mathbf{g}\left(\mathbf{x}^{(t+1)}\right) \end{aligned}$$

The minimization on $\mathbf{x}$ is done with the SciPy minimize. The code is provided in the notebook attached.

Results from numerical analysis confirm the results from analytical analysis. Aside from inaccuracies introduced by low value of $\mu_{01}(y_{01})$, the conditional distribution given Y_{01} matches the exponential distribution $\exp[\mathcal{H}(s)]$ to around 10^{-6} accuracy. Additionally, any distribution on Y_{01} can be optimal, which conforms with the analytical understanding.

1.4 Correlation

To see the correlation, we will try to calculate the Pearson coefficient for the pairs of random variables. We will test these on different choices of β and B. To make sure that the results are accurate, we reduced the accuracy requirement to below 10^{-10}.

- For Y_{01}^*, and X_0^*, we can see that the Pearson coefficient is very close to 0 except for when B is much larger than 1 and Y_{01}^* is skewed to one of the results, which could be just an accuracy issue. We hence hypothesize that Y_{01}^*, and X_0^* are independent. To make it more certain, we can try numerically calculating the distribution in exponential form to confirm.

It is unclear what correlation is like in this situation. Given how the distribution is structured, we can separate the exponent additively to obtain mutually independent pieces. For example, if we fix x_0, we know that X_1, Y_{01} and X_2, Y_{02} are independent of each other:

$$\mu(s) = \frac{1}{Z} \exp(Bx_0) \exp\left[\frac{\beta}{2}(x_0 x_1 y_{01}) + \alpha y_{01}\right] \exp\left[\frac{\beta}{2} x_0 x_2 y_{02}\right]$$

I have tried to calculate the covariance between different random variables from the numerical results, but they differ by the choice of μ_{01}, and the marginal distributions don't seem like there is a definitive answer for if they correlate or are independent. I may need more guidance on this issue.

2 Jun 6 - Jun 12, 2023

To simplify, we still consider the Ising model.

2.1 Numerical Analysis

2.1.1 Method

We vectorize the state, representing the probability as $\mu_{x_0 k}$, where x_0 is the spin state for the root node, and k leaves are in the spin-up state. The Hamiltonian can then be rewritten

$$\mathcal{H}(x_0, k) = \frac{\beta}{2} x_0(2k - \kappa) + Bx_0.$$

The edge distribution is therefore, for some distribution p,

$$\pi_p(x_0, x_v) = \frac{1}{\kappa} \sum_{k=1}^{\kappa} p_{x_0 k} \left[\mathbf{1}_{\{x_v = 1\}} k + \mathbf{1}_{\{x_v = -1\}} (\kappa - k) \right].$$

Note that if we represent $\mu \in \mathbb{R}^{2\times(\kappa+1)}$, this can be represented as a matrix multiplication.

The underlying distribution needs to account for multiplicity as well:

$$\eta_{x_0 k} = \binom{\kappa}{k} 2^{-\kappa}.$$

Hence, the optimization problem becomes

$$\begin{aligned} \text{maximize} \quad & \sum_{x_0 \in \mathcal{X}} \sum_{k=0}^{\kappa} \mu_{x_0 k} \mathcal{H}(x_0, k) - \sum_{x_0 \in \mathcal{X}} \sum_{k=0}^{\kappa} \mu_{x_0 k} \log \frac{\mu_{x_0 k}}{\eta_{x_0 k}} + \frac{\kappa}{2} \sum_{x_0 \in \mathcal{X}} \sum_{x_v \in \mathcal{X}} \pi_\mu(x_0, x_v) \log \frac{\pi_\mu(x_0, x_v)}{\pi_\eta(x_0, x_v)} \\ \text{subject to} \quad & \sum_{x_0 \in \mathcal{X}} \sum_{k=0}^{\kappa} \mu_{x_0 k} = 1 \\ & \pi_\mu(1, -1) = \pi_\mu(-1, 1). \end{aligned}$$

Like last time, to guarantee positivity, we instead optimize for some ξ such that $\mu_{x_0 k} = \exp(\xi_{x_0 k})$. And again, we optimize with a proximal point method. Proximal points are calculated with L-BFGS-B from SciPy. The tolerances of the optimization target, μ^*, and constraints are 10^{-6}.

2.1.2 Objects of interest

- First consider X_0 and X_1, where X_1 is a leaf chosen uniformly. The joint distribution is

$$\mathbb{P}_\mu(X_0 = x_0, X_1 = x_1) = \pi_\mu(x_0, x_1).$$

  Since $\pi_\mu(x_0, x_1)$ is symmetric, they are identically distributed. From numerical analysis, we find that only when β is 0 are X_0 and X_1 independent. When $\beta > 0$, there is a preference of X_0 and X_1 being the same spins, and vice versa. For $B = 0$, As $|\beta| \to \infty$, we see that X_0 and X_1 goes to Bernoulli(1/2).

- Then we consider X_1 and X_2, where they are distinct leaves chosen uniformly:

$$\mathbb{P}_\mu(X_1 = x_1, X_2 = x_2) = \frac{1}{\binom{\kappa}{k}} \sum_{x_0 \in \mathcal{X}} \mu_{x_0 k},$$

  where k is the number of spin-up states in $\{x_1, x_2\}$ Results from the simulation conclude that X_1 and X_2 are only independent when $\beta = 0$. The higher the β is, we approach $X_1 = X_2$.

- For X_1 and X_2 conditioned under X_0, we have

$$\mathbb{P}_\mu(X_1 = x_1, X_2 = x_2 | X_0 = x_0) = \frac{1}{\binom{\kappa}{k}} \mu_{x_0 k}.$$

  Again, X_1 and X_2 are only independent when $\beta = 0$. When $\beta > 0$, they tend to equal X_0, and when $\beta < 0$, they tend to equal $-X_0$, consistent with the first conclusion.

This points towards a representation of μ^* similar to the form

$$\mu^*(x_0, x_1, x_2) = \frac{1}{Z} \exp(\mathcal{H}(x_0, x_1, x_2)).$$