3 Exact Inference, Variational Inference and Gibbs Sampling

3.1 Message Passing

3.1.1 Description

Using techniques from Exercise 6.7 of the textbook, we first reduce the Markov network representation of the Ising model along the column and obtain a linear chain of nodes. We denote the set of all nodes along the same i-th column as $\gamma_i = \{x_{j,i} \mid j \in \{1,...,10\}\}$ for $i \in \{1,...,10\}$, and define the factors between the nodes as $f_i = f(\gamma_i,\ \gamma_{i+1})$ for $i \in \{1,...9\}$. An illustrated example of the reduced factor graph is shown below.

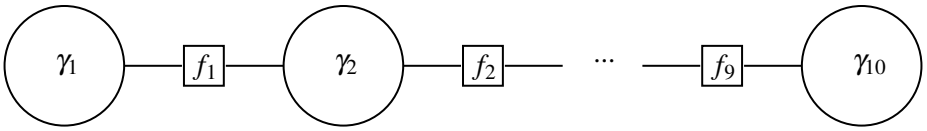

Figure 1: Induced factor-graph representation of the 10x10 Ising model.

More specifically, since $f_i = f(\gamma_i,\ \gamma_{i+1})$ represents a function between the two neighboring nodes, we define it as the product of both within-column and between-column potentials.

$$f(\gamma_i,\ \gamma_{i+1}) = \prod_{j=1}^{9} \phi(x_{j,i}, x_{j+1,i}) \prod_{j=1}^{9} \phi(x_{j,i+1}, x_{j+1,i+1}) \prod_{k=1}^{10} \phi(x_{k,i}, x_{k,i+1})$$

For example, the leftmost factor $f_1 = f(\gamma_1,\ \gamma_2)$ would look like:

$$f(\gamma_1,\ \gamma_2) = \prod_{j=1}^{9} \phi(x_{j,1}, x_{j+1,1}) \prod_{j=1}^{9} \phi(x_{j,2}, x_{j+1,2}) \prod_{k=1}^{10} \phi(x_{k,1}, x_{k,2})$$

Next, We perform exact inference on the factor graph from left to right:

$$\begin{aligned}
\mu_{\gamma_1 \to \gamma_2}(x) &= \sum_{\gamma_1} f(\gamma_1, \gamma_2) \\
\mu_{\gamma_2 \to \gamma_3}(x) &= \sum_{\gamma_2} f(\gamma_2, \gamma_3) \sum_{\gamma_1} f(\gamma_1, \gamma_2) \\
&\cdots \\
\mu_{\gamma_9 \to \gamma_{10}}(x) &= \sum_{\gamma_9} f(\gamma_9, \gamma_{10}) \sum_{\gamma_8} f(\gamma_8, \gamma_9) \sum_{\gamma_7} \cdots \sum_{\gamma_1} f(\gamma_1, \gamma_2)
\end{aligned}$$

This gives us the joint probability of the rightmost column, since the messages come only from one side:

$$P(\gamma_{10}) \propto \mu_{\gamma_9 \to \gamma_{10}}$$

which we can marginalize over all variables that are not the top and bottom nodes:

$$P(x_{1,10}, x_{10,10}) = \frac{1}{Z} \sum_{x_2,\ldots,x_9} \mu_{\gamma_9 \to \gamma_{10}}$$

3.1.2 Results

Message passing, conducted using three values of β, produces the following probability tables. Notably, this process effectively preserves long-distance dependencies between variables—a key characteristic of Ising models.

		$x_{10,10}$	
		0	1
$x_{1,10}$	0	0.499829	0.000171
	1	0.000171	0.499829

Table 1: $\beta = 4$

		$x_{10,10}$	
		0	1
$x_{1,10}$	0	0.412854	0.087146
	1	0.087146	0.412854

Table 2: $\beta = 1$

		$x_{10,10}$	
		0	1
$x_{1,10}$	0	0.25125	0.24875
	1	0.24875	0.25125

Table 3: $\beta = 0.01$

3.2 Variational Inference

3.2.1 Mean Field Approximation

We approximate the distribution of the Ising lattice with a simpler factor-able model:

$$Q(X) = \prod_{x_i} q_i(x_i),\ q_i(x_i) \sim Bernoulli(\mu_i)$$

In other words, we model each variable as an individually distributed binary Bernoulli distribution that takes two values: 0 and 1. More specifically, for each variable x_i we define an associated parameter μ_i , such that $\mu_i = P(x_i = 1)$.

3.2.2 Coordinate Ascent

At each iteration of the coordinate ascent algorithm, we choose a variable x_k and optimize its distribution $q_k(x_k)$, while fixing the distribution of all other variables. It is shown via the correlation of ELBO and KL-divergence that for each update $q_k^*(x_k)$, we have the following relationship:

$$q_k^*(x_k) \propto e^{<logp(x_k|x_{-k})>_{q_{-k}}}$$

For the Ising model given, it is trivial to show that

$$\begin{aligned} p(x_k|x_{-k}) &= \frac{e^{\beta \sum_{j\in\mathcal{N}(k)} \mathbf{I}[x_j=x_k]}}{e^{\beta \sum_{j\in\mathcal{N}(k)} \mathbf{I}[x_j=1]} + e^{\beta \sum_{j\in\mathcal{N}(k)} \mathbf{I}[x_j=0]}} \\ &= e^{\beta \sum_{j\in\mathcal{N}(k)} \mathbf{I}[x_j=x_k]} \cdot C \end{aligned} \tag{1}$$

for some constant C w.r.t states of x_k. Thus, we obtain that

$$\begin{aligned} q_k^*(x_k) &\propto e^{<\beta \sum_{j\in\mathcal{N}(k)} \mathbf{I}[x_j=x_k]>_{q_{-k}}} \\ &= e^{\beta \sum_{j\in\mathcal{N}(k)} \sum_{x_{-k}} q_{-k}(x_{-k}) \mathbf{I}[x_j=x_k]} \\ &= e^{\beta \sum_{j\in\mathcal{N}(k)} \sum_{x_j} q_j(x_j) \mathbf{I}[x_j=x_k] \cdot \sum_{x_{-k}\setminus x_j} \prod_{i\neq j, i\neq k} q_i(x_i)} \\ &= e^{\beta \sum_{j\in\mathcal{N}(k)} \sum_{x_j} q_j(x_j) \mathbf{I}[x_j=x_k] \cdot 1} \\ \therefore q_k^*(x_k) &\propto e^{\beta \sum_{j\in\mathcal{N}(k)} (\mu_j \mathbf{I}[x_k=1] + (1-\mu_j)\mathbf{I}[x_k=0])} \end{aligned} \tag{2}$$

Finally, we get to use (2) as the update equation in our Python implementation.

3.2.3 Results

Running coordinate ascent with the mean-field approximation using three different values of β produces the following probability tables. Notably, when β is high, the distribution exhibits significant bias. This occurs because large β values imply strong coupling between the variables—a dependency entirely overlooked by the independence assumption of the mean-field approximation."

		$x_{10,10}$	
		0	1
$x_{1,10}$	0	0.999329	0.000335
	1	0.000335	0.000000

Table 4: $\beta = 4$

		$x_{10,10}$	
		0	1
$x_{1,10}$	0	0.714738	0.130684
	1	0.130684	0.023895

Table 5: $\beta = 1$

		$x_{10,10}$	
		0	1
$x_{1,10}$	0	0.25	0.25
	1	0.25	0.25

Table 6: $\beta = 0.01$

3.3 Gibbs Sampling

3.3.1 Method

Our implementation of Gibbs sampling chooses a variable at random in each iteration for update. We define our transition probability (update function) as:

$$T(x \to x') = \frac{1}{N} p(x'_i | x_{-i})$$

where x'_i is the variable to update, while x_{-i} is the state of all other variables in the previous round. By definition of the Ising model, $p(x'_i|x_{-i})$ is equation (1).

3.3.2 Results

Running Gibbs sampling with three different values of β produces the following probability tables. Each table corresponds to a single run of the sampling procedure, which accounts for the skewness observed due to the method's local nature. This highlights both the tendency of Gibbs sampling to converge to local optima and its sensitivity to initialization.

		$x_{10,10}$	
		0	1
$x_{1,10}$	0	0.997739	0.001487
	1	0.000756	0.000018

Table 7: $\beta = 4$

		$x_{10,10}$	
		0	1
$x_{1,10}$	0	0.354386	0.232140
	1	0.205180	0.208294

Table 8: $\beta = 1$

		$x_{10,10}$	
		0	1
$x_{1,10}$	0	0.257205	0.250378
	1	0.245429	0.246988

Table 9: $\beta = 0.01$