Advanced Machine Week 5

Benedetta Felici, Jasper Pieterse & Daria Mihaila

January 2024

Mean Field Approximation and Belief Propagation for the Ising Model

In this report, we conduct a comparative analysis of two methodologies for obtaining approximate solutions of mean spin values and correlations in the Ising model: the mean field approximation and the belief propagation algorithm.

Our investigation spans a diverse range of experimental setups, encompassing various instances of the Ising model distinguished by differing spin connectivities and sparsity levels.

More specifically, we answer the following research questions:

- What is the accuracy of the procedures in approximating the mean spin values and correlations for the Fully Connected Ising model?
- How do the procedures fare in estimating the mean spin values for the Sparse Ising model?
- Is one procedure more efficient than the other, considering the number of iterations required for convergence?
- How do the accuracy and efficiency of the approximations compare with different parameters of the Ising model?

Following in the report, we give an overview of the underlying theory behind the exact and approximate solutions of the Ising model, discussing in detail their implementation. Finally, we report the results of the simulations alongside a comprehensive reflection.

Theory

Ising Model

The Ising model is defined by the probability distribution:

$$p(x) = \frac{1}{Z} exp(\sum_{i,j} w_{ij} x_i x_j + \sum_i \theta_i x_i) \tag{1}$$

In the expression above, the w_{ij} denote the coupling between spins, which follow a Gaussian distribution with mean $\frac{J_0}{k}$ and variance $\frac{J^2}{k}$ where $k = cn$. The θ_i values, representing an external magnetic field, are either kept fixed or randomly distributed.

Within the model, the exact correlations between spins are defined as:

$$\chi_{ij}^{ex} = < s_i s_j > - m_i m_j \tag{2}$$

where m_i and m_j denote the expected value of the corresponding spin.

Mean Field Approximation

The mean field (MF) approximation consists of a variational approximation of the probability distribution in Equation 1. This yields a probability distribution $q(x)$ which is found by minimizing the KL divergence with $p(x)$.

The distribution q can be expressed in terms of $m_i =< x_i >_q$, denoting the expectation over q of spin i. It can be shown analytically that the mean field approximation corresponds to the solution of the following set of equations:

$$m_i = \sum_{j \neq i} tanh(w_{ij} m_j + \theta_i) \tag{3}$$

Within the mean field approximation, it is possible to compute easily the correlations between spins χ_{ij}^{MF} by derivating the logarithm the partition function Z with respect to the θ parameters. This entails the following expression for the correlations:

$$\begin{aligned} \chi_{ij}^{MF} &= A_{ij}^{-1} \\ A_{ij} &= \frac{\delta_{ij}}{1 - m_i^2} - w_{ij} \end{aligned} \tag{4}$$

where δ_{ij} is the Kronecker delta function computed in i and j.

Belief Propagation

The belief propagation (BP) is an iterative message passing algorithm over a given graphical model with the goal to perform inference.

In this framework, the mean spins $m_{ij}(x_j)$ can be interpreted as a message from node i to j. This can be estimated as

$$m_{ij}(x_j) = \sum_{x_i} \psi_{ij}(x_i, x_j) \psi_i(x_i) \prod_{k \in N(i) - j} m_{ki}(x_i) \tag{5}$$

We can parametrize the message as $m_{ij}(x_j) = exp(a_{ij} x_j)$. Therefore, equation 5 can be written as:

$$m_{ij}(x_j) = 2cosh(w_{ij} x_j + \theta_i + \sum_{k \in N(i) - j} a_{ki}) \tag{6}$$

where the a_{ij} can be computed as

$$a_{ij} = \frac{1}{2} log \frac{m_{ij}(x_j = 1)}{m_{ij}(x_j = 1)} \tag{7}$$

By combining equations 6 and 7, it is possible to write an iterative algorithm which converges to the solutions of the message passing equations $m_{ij}(x_j)$. After convergence, the correlations between spins can be computed as:

$$\chi_{ij}^{BP} = \sum_{x_i, x_j} x_i x_j b_{ij}(x_i, x_j) - m_i^{BP} m_j^{BP} \tag{8}$$

where b_{ij} denotes the BP approximation of the joint probability distribution of x_i and x_j.

Experimental Setup

In this report, we focus on understanding the impact of several parameters of the Ising model on the accuracy and efficiency of the MF and BP approximations. More specifically, we vary β, which impacts the strength of the connectivities between spins, θ, which denotes an external magnetic field and c, which determines the sparsity of the networks. Following in this Section, we first provide details on how instances of the Ising model are generated. Then, we describe in detail the procedure to obtain the BP and MF approximations. Finally, we explain in detail how we visualize and compare the results.

Ising Model Generation

Firstly, we initialize an instance of the Ising model with dimensionality n=20. Two distinct configurations are considered: the Fully Connected Ising model and the Sparse Ising model.

To obtain an Fully Connected instance of the Ising model, we set $c = 1$ and sample the weights from a Gaussian distribution, as explained beforehand. We set $J_0 = 0$ and $J = \beta$, with varying levels of β between 0 and 1. The θ_i are also generated randomly with a Gaussian, and multiplied by β.

On the other hand, to obtain a Sparse Ising model we set the weights to β, with sign determined from the value generated by the Gaussian distribution. In this case, we do not generate the θ randomly but we set them to a fixed value $\theta > 0$. In this case, we also investigate the impact of c in the simulations, which we vary between 0 and 1. Indeed, the parameter c regulates the sparsity of the network. When $c = 1$, we have a fully connected Ising model, and decreasing c results in progressively sparser networks, with approximately cn^2 non-zero elements in w.

Mean Field Approximation Implementation Details

We compute the mean field approximation m^{MF} by using the fixed point iteration method, which is described in Algorithm 1. The underlying idea behind the method is that by starting from a random m we obtain a progressively better approximation of the solution of Equation 3 by updating m with the weighted sum of m and the result of applying 3 to m. More specifically, the weights are determined by the smoothing parameter μ, set to 0.5 in our simulations. Furthermore, to prevent looping, we terminate the procedure once the difference between the old value and the new one is smaller than the precision parameter $\epsilon = 1e - 13$.

Algorithm 1 Mean Field Approximation

```
Initialize a random vector m
dm = 1
μ = 0.5
ϵ = 1 × 10^-13
while dm > ϵ do
    m_old = m
    Compute m = μm + (1 − μ) tanh( Σ_{j∈N(i)} w_ij m_j + θ_i )
    dm = max(|m − m_old|)
end while
The mean field approximation is m^MF = m
```

Belief Propagation Implementation Details

The BP solution is also computed as a fixed point iteration in the matrix A, previously defined in Equation 7. The procedure is described in detail in Algorithm 2. Also in this case, we use $\epsilon = 1e - 13$ and apply smoothing with $\mu = 0.5$ only to the sparse case.

Algorithm 2 Belief Propagation Algorithm

Start with a random initial $n \times n$ matrix A
$dA = 1$
$\epsilon = 1 \times 10^{-13}$
$iter = 0$
$N_{iter} = 1000$
while $dA > \epsilon$ and $iter < N_{iter}$ **do**
    $A_{\text{old}} = A$
    Compute matrix $m_{ij}(x_j = 1)$ and $m_{ij}(x_j = -1)$ according to Eq. 6
    Compute matrix A according to Eq. 7
    $dA = \max(\max(|A - A_{\text{old}}|))$
end while
$m_i^{BP} = \tanh(\theta_i + \sum_j a_{ji})$

During our simulations, we observed that the Belief Propagation algorithm exhibited failure in converging due to the errors between consecutive parameters exceeding the precision parameter repeatedly. To prevent potential looping, we implemented an additional stopping condition in Algorithm 2, terminating it after 1000 iterations if convergence doesn't occur.

Algorithmic Comparison

In all simulations, the root mean squared errors (RMSE) are calculated between the exact mean spins and those obtained through MF and BP approximations. We run the simulations for $N = 10$ numbers of different instances for all the different combinations of parameters we employ, and average the errors. The mean errors, alongside a confidence interval given by the standard deviation, are then plotted against β for the Fully Connected model, and against β and c for the Sparse model.

For the Fully Connected model, RMSE are also computed between the approximate and exact correlations, and plotted against β.

For the Sparse model, we conduct simulations by varying values of the parameters c, β, and θ. More specifically, we plot the dependency of the error and iterations against c where $\beta = 0.2$, and against β where $c = 0.2$, for different values of $\theta = 0.2, 0.5, 0.8$.

Results

Fully connected Ising Model

In the fully connected case, we see from Figure 1 that the BP algorithm performs better than the MF for all values of β, as it results in diminuished errors both in the approximations of the mean spins and the correlations. Both the procedure terminate in several hundred iterations, with the BP being substantially faster than the MF.

The results also indicate that the approximations of the mean spins and the correlations for both the BP and the MF are much more precise with lower levels of the parameter β, and the procedures are substantially faster as well.

Furthermore, both the MF and BP approximation exhibit substantial accuracy variation depending on the considered instance of the Ising model, as visible from the large standard deviation, especially for higher values of β.

The results are in line with the ones provided in the assignment guidelines, with the exception of the number of iterations of the BP model being substantially larger than the MF for high values of the parameter β. Indeed, as previously mentioned, the BP procedure failed

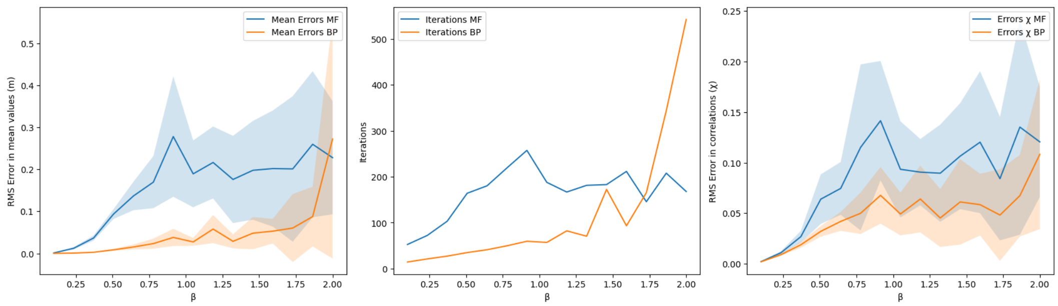


Figure 1: Simulation results for Mean Field (MF) and Belief Propagation (BP) in a fully connected Ising Model. Left: RMSE between approximate and exact mean spin values, with background color indicating a confidence interval given by the standard deviation. Center: Mean of iterations until convergence. Right: RMSE in exact and approximate correlations, with standard deviation. All metrics are plotted against the connectivity parameter β

to converge in our simulations multiple times, depending on the specific instance of Ising model. Therefore, it is possible that this didn't occur for the instance of the Ising model generated in the assignment.

Sparse Ising Model

Overall, we observe that for sparse networks the BP algorithm generally results in better approximations.

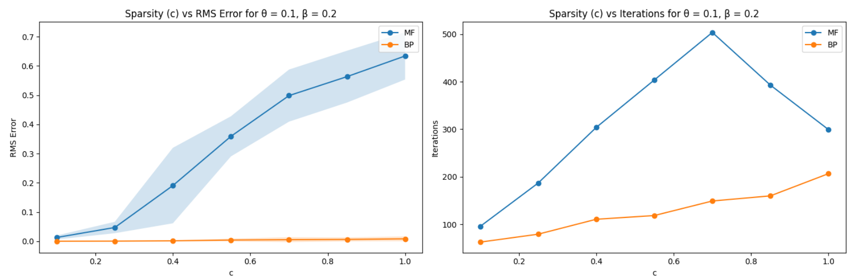
More specifically, in Figure 3 we find that for sparse networks ($c = 0.2$) the MF errors are higher than the BP for almost all levels of couplings, except very small ones ($\beta = 0.1$) for which the algorithms performs comparably. Similarly, in Figure 2, it is shown that for networks with small couplings ($\beta = 0.2$) BP yields a better approximation than MF for every level of sparsity considered. This is in line with what we expected, with the exception that the BP procedure yields a greater performance than the one reported in the assignment guidelines, particularly for small values of β, as visible in Figure 2. However, we think this is in line with the theory, given that the BP approximation on tree-like networks is exact, and the more the network is sparse, the less we are likely to incur into a graph with loops and multiple paths between two nodes.

In terms of efficiency, for smaller values of c and β the BP seems to perform better than the MF, however the BP fails to converge especially with high values of β, as visible in Figure 3. Although a similar outcome was not reported in the assignment, we think this behaviour is expected, as there is no theoretical guarantee for the BP procedure to converge.

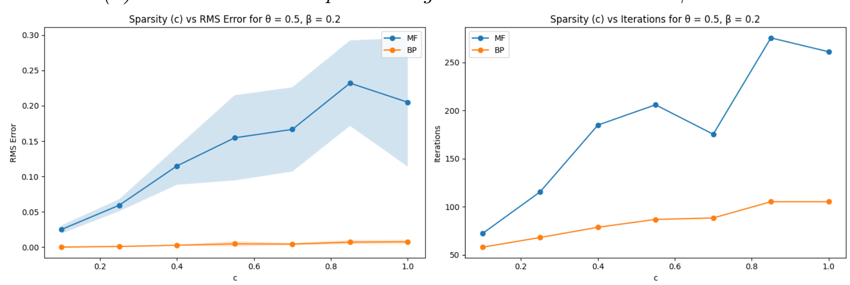
In terms of dependency over θ, the performance of the two algorithms, especially the MF, was found to be substantially better with larger values of this parameter.

Conclusion

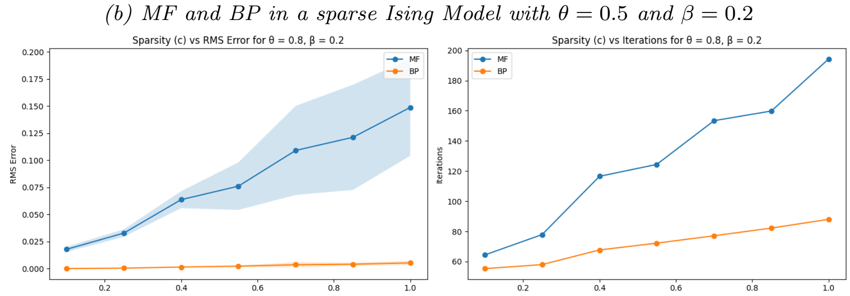
Overall, we implemented and compared two algorithms with the goal to perform approximate inference of the Ising model. To summarize, we found the BP algorithm to generally yield better approximations and to converge faster than the MF. However, the results showed great variability depending on both the parameters which characterize the generated instance of the Ising model, and the specific instance itself.



(a) *MF and BP in a sparse Ising Model with $\theta = 0.1$ and $\beta = 0.2$*

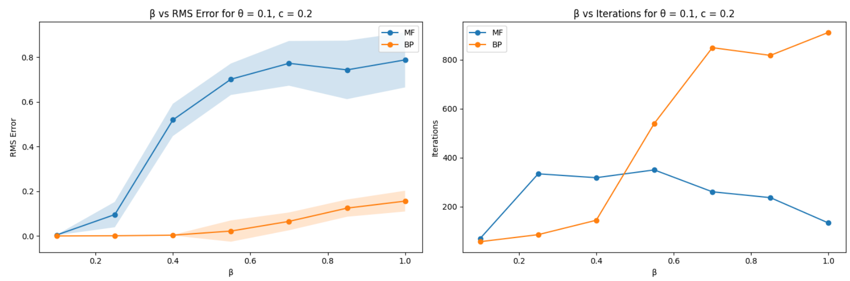


(b) *MF and BP in a sparse Ising Model with $\theta = 0.5$ and $\beta = 0.2$*

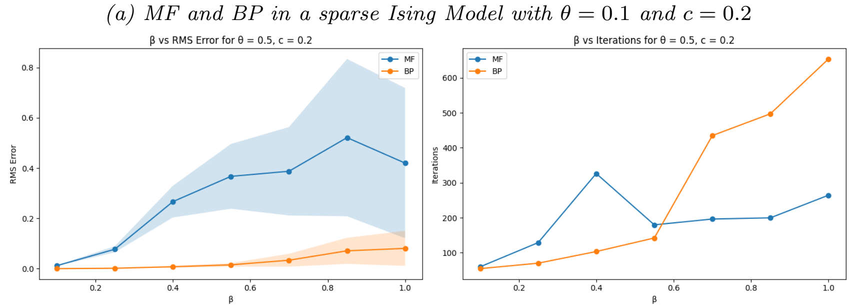


(c) *MF and BP in a sparse Ising Model with $\theta = 0.8$ and $\beta = 0.2$*

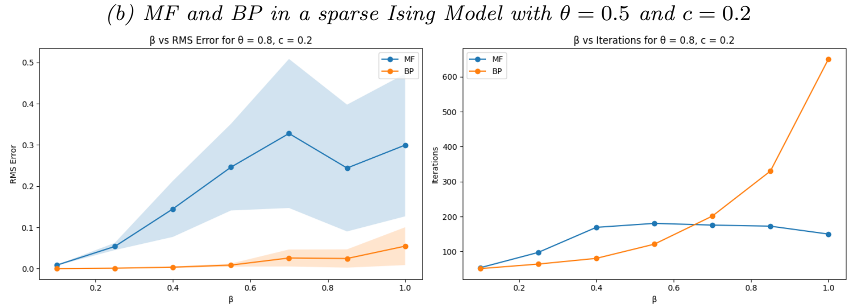
Figure 2: Simulation results on the effect of $0 < c \leq 1$ on the inference of the Sparse Ising model



(a) MF and BP in a sparse Ising Model with $\theta = 0.1$ and $c = 0.2$



(b) MF and BP in a sparse Ising Model with $\theta = 0.5$ and $c = 0.2$



(c) MF and BP in a sparse Ising Model with $\theta = 0.8$ and $c = 0.2$

Figure 3: Simulation results on the effect of $0 < \beta \leq 1$ on the inference of the Sparse Ising model