Mean field Ising models

Sumit Mukherjee

Columbia University

Joint work with Anirban Basak, Duke University

What is an Ising model?

- An Ising model is a probability distribution on $\{-1, 1\}^n$.

What is an Ising model?

- An Ising model is a probability distribution on $\{-1, 1\}^n$.
- The probability mass function of the Ising model is given by

$$\mathbb{P}_n(X = \mathbf{x}) = e^{\frac{\beta}{2}\mathbf{x}' A_n \mathbf{x} - F_n}.$$

What is an Ising model?

- An Ising model is a probability distribution on $\{-1, 1\}^n$.
- The probability mass function of the Ising model is given by

$$\mathbb{P}_n(X = \mathbf{x}) = e^{\frac{\beta}{2}\mathbf{x}' A_n \mathbf{x} - F_n}.$$

- Here A_n is a symmetric $n \times n$ matrix with 0 on the diagonals, and $\beta > 0$ is a one dimensional parameter usually referred to as inverse temperature.

What is an Ising model?

- An Ising model is a probability distribution on $\{-1, 1\}^n$.
- The probability mass function of the Ising model is given by

$$\mathbb{P}_n(X = \mathbf{x}) = e^{\frac{\beta}{2}\mathbf{x}' A_n \mathbf{x} - F_n}.$$

- Here A_n is a symmetric $n \times n$ matrix with 0 on the diagonals, and $\beta > 0$ is a one dimensional parameter usually referred to as inverse temperature.
- F_n is the log partition function, which will be the main focus of this talk.

Applications of Ising model

- *Ferromagnetism*: Ising model was introduced by Wilhelm Lenz in 1920 to study orientation of configurations of electrons in Ferromagnetic materials. The idea is that if most of the electrons are oriented in the same direction, we observe magnetism.

Applications of Ising model

- *Ferromagnetism*: Ising model was introduced by Wilhelm Lenz in 1920 to study orientation of configurations of electrons in Ferromagnetic materials. The idea is that if most of the electrons are oriented in the same direction, we observe magnetism.
- *Lattice gas*: Ising models have been used to model motion of atoms of gas, by partitioning space into a lattice. Each position is either 1 or -1 depending on whether an atom is present or absent.

Applications of Ising model

- *Ferromagnetism*: Ising model was introduced by Wilhelm Lenz in 1920 to study orientation of configurations of electrons in Ferromagnetic materials. The idea is that if most of the electrons are oriented in the same direction, we observe magnetism.
- *Lattice gas*: Ising models have been used to model motion of atoms of gas, by partitioning space into a lattice. Each position is either 1 or -1 depending on whether an atom is present or absent.
- *Neuroscience*: Ising models have been used to study neuron activity in our brains. Here each neuron is $+1$ if active, and -1 if inactive.

Applications of Ising model

- *Ferromagnetism*: Ising model was introduced by Wilhelm Lenz in 1920 to study orientation of configurations of electrons in Ferromagnetic materials. The idea is that if most of the electrons are oriented in the same direction, we observe magnetism.
- *Lattice gas*: Ising models have been used to model motion of atoms of gas, by partitioning space into a lattice. Each position is either 1 or -1 depending on whether an atom is present or absent.
- *Neuroscience*: Ising models have been used to study neuron activity in our brains. Here each neuron is $+1$ if active, and -1 if inactive.
- A generalization of the Ising model for more than two states (Potts model) has been used to study Protein folding and Image Processing.

Commonly studied Ising models

Commonly studied Ising models

- *Ferromagnetic Ising model* on generals graphs: A_n is the adjacency matrix of a finite graph (scaled by the average degree):

Commonly studied Ising models

- *Ferromagnetic Ising model* on generals graphs: A_n is the adjacency matrix of a finite graph (scaled by the average degree):
  - Lattice graphs: Original paper by Ising (1925) considered the one dimensional lattice, subsequent focus has been on higher dimensional lattices;

Commonly studied Ising models

- *Ferromagnetic Ising model* on generals graphs: A_n is the adjacency matrix of a finite graph (scaled by the average degree):
  - Lattice graphs: Original paper by Ising (1925) considered the one dimensional lattice, subsequent focus has been on higher dimensional lattices;
  - Curie-Weiss Model: Complete graph;

Commonly studied Ising models

- *Ferromagnetic Ising model* on generals graphs: A_n is the adjacency matrix of a finite graph (scaled by the average degree):
  - Lattice graphs: Original paper by Ising (1925) considered the one dimensional lattice, subsequent focus has been on higher dimensional lattices;
  - Curie-Weiss Model: Complete graph;
  - Random Graphs: Erdős-Rényi, random regular, etc.

Commonly studied Ising models

- *Ferromagnetic Ising model* on generals graphs: A_n is the adjacency matrix of a finite graph (scaled by the average degree):
  - Lattice graphs: Original paper by Ising (1925) considered the one dimensional lattice, subsequent focus has been on higher dimensional lattices;
  - Curie-Weiss Model: Complete graph;
  - Random Graphs: Erdős-Rényi, random regular, etc.
- *Sherrington-Kirkpatrick*: A_n is a matrix of i.i.d. $\mathcal{N}(0, 1/n)$.

Commonly studied Ising models

- *Ferromagnetic Ising model* on generals graphs: A_n is the adjacency matrix of a finite graph (scaled by the average degree):
  - Lattice graphs: Original paper by Ising (1925) considered the one dimensional lattice, subsequent focus has been on higher dimensional lattices;
  - Curie-Weiss Model: Complete graph;
  - Random Graphs: Erdős-Rényi, random regular, etc.
- *Sherrington-Kirkpatrick*: A_n is a matrix of i.i.d. $\mathcal{N}(0, 1/n)$.
- *Hopfield model* of neural networks: where $A_n = \frac{1}{n} B_n B_n'$, with B_n a matrix of i.i.d. random matrix taking values ± 1 with probability $\frac{1}{2}$.

Why is the log partition function important?

- The log partition function gives information about the moments of the distribution. For e.g.

$$\frac{\partial F_n}{\partial A_n(i,j)} = \mathbb{E}(X_i X_j).$$

Similar statements holds for higher derivatives.

Why is the log partition function important?

- The log partition function gives information about the moments of the distribution. For e.g.

$$\frac{\partial F_n}{\partial A_n(i,j)} = \mathbb{E}(X_i X_j).$$

  Similar statements holds for higher derivatives.

- The partition function gives information about the stable configurations under the model.

Why is the log partition function important?

- The log partition function gives information about the moments of the distribution. For e.g.

$$\frac{\partial F_n}{\partial A_n(i,j)} = \mathbb{E}(X_i X_j).$$

  Similar statements holds for higher derivatives.

- The partition function gives information about the stable configurations under the model.

- The uniform distribution on the space of q colorings of a graph is a (limiting) Potts model, and the partition function is the number of q colorings of the graph.

Why is the log partition function important?

- The log partition function gives information about the moments of the distribution. For e.g.

$$\frac{\partial F_n}{\partial A_n(i,j)} = \mathbb{E}(X_i X_j).$$

  Similar statements holds for higher derivatives.

- The partition function gives information about the stable configurations under the model.

- The uniform distribution on the space of q colorings of a graph is a (limiting) Potts model, and the partition function is the number of q colorings of the graph.

- Many estimation techniques in Statistics require the knowledge, or good estimates of the partition function.

Mean field technique of Statistical Physics

- Let $D(\cdot||\cdot)$ denote the Kullback-Leibler divergence.

Mean field technique of Statistical Physics

- Let $D(\cdot||\cdot)$ denote the Kullback-Leibler divergence.
- Then for any probability measure $\mathbb{Q}_n$ on $\{-1,1\}^n$, $D(\mathbb{Q}_n||\mathbb{P}_n)$ equals

$$\sum_{\mathbf{x}\in\{-1,1\}^n} \mathbb{Q}_n(\mathbf{x}) \log \frac{\mathbb{Q}_n(\mathbf{x})}{\mathbb{P}_n(\mathbf{x})}$$

Mean field technique of Statistical Physics

- Let $D(\cdot||\cdot)$ denote the Kullback-Leibler divergence.
- Then for any probability measure $\mathbb{Q}_n$ on $\{-1,1\}^n$, $D(\mathbb{Q}_n||\mathbb{P}_n)$ equals

$$\sum_{\mathbf{x}\in\{-1,1\}^n} \mathbb{Q}_n(\mathbf{x}) \log \frac{\mathbb{Q}_n(\mathbf{x})}{\mathbb{P}_n(\mathbf{x})}$$

$$= \sum_{\mathbf{x}\in\{-1,1\}^n} \mathbb{Q}_n(\mathbf{x}) \log \mathbb{Q}_n(\mathbf{x}) - \sum_{\mathbf{x}\in\{-1,1\}^n} \mathbb{Q}_n \log \mathbb{P}_n(\mathbf{x})$$

Mean field technique of Statistical Physics

- Let $D(\cdot||\cdot)$ denote the Kullback-Leibler divergence.
- Then for any probability measure $\mathbb{Q}_n$ on $\{-1,1\}^n$, $D(\mathbb{Q}_n||\mathbb{P}_n)$ equals

$$\begin{aligned}
&\sum_{\mathbf{x}\in\{-1,1\}^n} \mathbb{Q}_n(\mathbf{x}) \log \frac{\mathbb{Q}_n(\mathbf{x})}{\mathbb{P}_n(\mathbf{x})} \\
&= \sum_{\mathbf{x}\in\{-1,1\}^n} \mathbb{Q}_n(\mathbf{x}) \log \mathbb{Q}_n(\mathbf{x}) - \sum_{\mathbf{x}\in\{-1,1\}^n} \mathbb{Q}_n \log \mathbb{P}_n(\mathbf{x}) \\
&= \sum_{\mathbf{x}\in\{-1,1\}^n} \mathbb{Q}_n(\mathbf{x}) \log \mathbb{Q}_n(\mathbf{x}) - \frac{\beta}{2} \sum_{\mathbf{x}\in\{-1,1\}^n} \mathbb{Q}_n(x) \mathbf{x}' A_n \mathbf{x} \\
&+ \sum_{\mathbf{x}\in\{-1,1\}^n} \mathbb{Q}_n(\mathbf{x}) F_n
\end{aligned}$$

Mean field technique of Statistical Physics

- Let $D(\cdot||\cdot)$ denote the Kullback-Leibler divergence.
- Then for any probability measure $\mathbb{Q}_n$ on $\{-1,1\}^n$, $D(\mathbb{Q}_n||\mathbb{P}_n)$ equals

$$\begin{aligned}
&\sum_{\mathbf{x}\in\{-1,1\}^n} \mathbb{Q}_n(\mathbf{x}) \log \frac{\mathbb{Q}_n(\mathbf{x})}{\mathbb{P}_n(\mathbf{x})} \\
=&\sum_{\mathbf{x}\in\{-1,1\}^n} \mathbb{Q}_n(\mathbf{x}) \log \mathbb{Q}_n(\mathbf{x}) - \sum_{\mathbf{x}\in\{-1,1\}^n} \mathbb{Q}_n \log \mathbb{P}_n(\mathbf{x}) \\
=&\sum_{\mathbf{x}\in\{-1,1\}^n} \mathbb{Q}_n(\mathbf{x}) \log \mathbb{Q}_n(\mathbf{x}) - \frac{\beta}{2} \sum_{\mathbf{x}\in\{-1,1\}^n} \mathbb{Q}_n(x)\mathbf{x}' A_n \mathbf{x} \\
+&\sum_{\mathbf{x}\in\{-1,1\}^n} \mathbb{Q}_n(\mathbf{x}) F_n \\
=&\mathbb{E}_{\mathbb{Q}_n}\left\{ \log \mathbb{Q}_n(X) - \frac{\beta}{2} X' A_n X \right\} + F_n.
\end{aligned}$$

- Since $D(Q_n||P_n) \geq 0$ with equality iff $\mathbb{Q}_n = \mathbb{P}_n$, we have

$$F_n \geq \mathbb{E}_{\mathbb{Q}_n}\left\{\frac{\beta}{2}X'A_nX - \log \mathbb{Q}_n(X)\right\},$$

with equality iff $\mathbb{Q}_n = \mathbb{P}_n$.

- Since $D(Q_n||P_n) \geq 0$ with equality iff $\mathbb{Q}_n = \mathbb{P}_n$, we have

$$F_n \geq \mathbb{E}_{\mathbb{Q}_n}\left\{\frac{\beta}{2}X'A_nX - \log \mathbb{Q}_n(X)\right\},$$

with equality iff $\mathbb{Q}_n = \mathbb{P}_n$.

- Equivalently,

$$F_n = \sup_{\mathbb{Q}_n} \mathbb{E}_{\mathbb{Q}_n}\left\{\frac{\beta}{2}X'A_nX - \log \mathbb{Q}_n(X)\right\},$$

where the sup is over all probability measures on $\{-1, 1\}^n$.

Mean field technique of Statistical Physics

- Instead, if we restrict the sup over product measures $\mathbb{Q}_n$, we get a lower bound

$$F_n \geq \sup_{\mathbb{Q}_n} \mathbb{E}_{\mathbb{Q}_n} \left\{ \frac{\beta}{2} X' A_n X - \log \mathbb{Q}_n(X) \right\}. \tag{1}$$

Mean field technique of Statistical Physics

- Instead, if we restrict the sup over product measures $\mathbb{Q}_n$, we get a lower bound

$$F_n \geq \sup_{\mathbb{Q}_n} \mathbb{E}_{\mathbb{Q}_n} \left\{ \frac{\beta}{2} X' A_n X - \log \mathbb{Q}_n(X) \right\}. \tag{1}$$

- Under a product measure $\mathbb{Q}_n$, each X_i is independently distributed with mean $m_i \in [-1, 1]$, say.

Mean field technique of Statistical Physics

- Instead, if we restrict the sup over product measures $\mathbb{Q}_n$, we get a lower bound

$$F_n \geq \sup_{\mathbb{Q}_n} \mathbb{E}_{\mathbb{Q}_n} \left\{ \frac{\beta}{2} X' A_n X - \log \mathbb{Q}_n(X) \right\}. \tag{1}$$

- Under a product measure $\mathbb{Q}_n$, each X_i is independently distributed with mean $m_i \in [-1, 1]$, say.

- Then X_i takes the value $\{1, -1\}$ with probabilities $\frac{1+m_i}{2}$ and $\frac{1-m_i}{2}$ respectively.

Mean field technique of Statistical Physics

- Instead, if we restrict the sup over product measures $\mathbb{Q}_n$, we get a lower bound

$$F_n \geq \sup_{\mathbb{Q}_n} \mathbb{E}_{\mathbb{Q}_n} \left\{ \frac{\beta}{2} X' A_n X - \log \mathbb{Q}_n(X) \right\}. \tag{1}$$

- Under a product measure $\mathbb{Q}_n$, each X_i is independently distributed with mean $m_i \in [-1, 1]$, say.

- Then X_i takes the value $\{1, -1\}$ with probabilities $\frac{1+m_i}{2}$ and $\frac{1-m_i}{2}$ respectively.

- Also $\mathbb{Q}_n(\mathbf{x})$ becomes $\prod_{i=1}^n \mathbb{Q}_{n,i}(x_i)$, which on taking expectation gives

$$\mathbb{E}_{\mathbb{Q}_n} \log \mathbb{Q}_n(X) = \sum_{i=1}^n \mathbb{E} \log \mathbb{Q}_{n,i}(X_i)$$

$$= \sum_{i=1}^{n} \Big[\frac{1+m_i}{2} \log \frac{1+m_i}{2} + \frac{1-m_i}{2} \log \frac{1-m_i}{2} \Big]$$

$$= \sum_{i=1}^{n} \Big[\frac{1+m_i}{2} \log \frac{1+m_i}{2} + \frac{1-m_i}{2} \log \frac{1-m_i}{2} \Big]$$

- Thus the RHS of (1) on the last slide becomes

$$\frac{\beta}{2} \mathbf{m}' A_n \mathbf{m} + \sum_{i=1}^{n} I(m_i),$$

where $I(x) = -\frac{1+x}{2} \log \frac{1+x}{2} - \frac{1-x}{2} \log \frac{1-x}{2}$ is the Binary entropy function.

$$= \sum_{i=1}^{n} \Big[\frac{1+m_i}{2} \log \frac{1+m_i}{2} + \frac{1-m_i}{2} \log \frac{1-m_i}{2} \Big]$$

- Thus the RHS of (1) on the last slide becomes

$$\frac{\beta}{2} \mathbf{m}' A_n \mathbf{m} + \sum_{i=1}^{n} I(m_i),$$

where $I(x) = -\frac{1+x}{2} \log \frac{1+x}{2} - \frac{1-x}{2} \log \frac{1-x}{2}$ is the Binary entropy function.

- Combining we have $F_n \geq M_n$, where $M_n = \sup_{\mathbf{m} \in [-1,1]^n} \left\{ \frac{\beta}{2} \mathbf{m}' A_n \mathbf{m} + \sum_{i=1}^{n} I(m_i) \right\}$.

$$= \sum_{i=1}^{n} \Big[\frac{1+m_i}{2} \log \frac{1+m_i}{2} + \frac{1-m_i}{2} \log \frac{1-m_i}{2} \Big]$$

- Thus the RHS of (1) on the last slide becomes

$$\frac{\beta}{2} \mathbf{m}' A_n \mathbf{m} + \sum_{i=1}^{n} I(m_i),$$

where $I(x) = -\frac{1+x}{2} \log \frac{1+x}{2} - \frac{1-x}{2} \log \frac{1-x}{2}$ is the Binary entropy function.

- Combining we have $F_n \geq M_n$, where $M_n = \sup_{\mathbf{m} \in [-1,1]^n} \Big\{ \frac{\beta}{2} \mathbf{m}' A_n \mathbf{m} + \sum_{i=1}^{n} I(m_i) \Big\}$.
- This lower bound to the log partition function is referred to as mean field prediction.

Eg: Mean field prediction for Regular graphs

- For a regular graph with degree d_n the matrix $A_n(i,j)$ is taken to be $\frac{1}{d_n}$ if $i \sim j$, and 0 otherwise.

Eg: Mean field prediction for Regular graphs

- For a regular graph with degree d_n the matrix $A_n(i,j)$ is taken to be $\frac{1}{d_n}$ if $i \sim j$, and 0 otherwise.
- This scaling ensures that the resulting Ising model is non trivial, for all values of d_n.

Eg: Mean field prediction for Regular graphs

- For a regular graph with degree d_n the matrix $A_n(i,j)$ is taken to be $\frac{1}{d_n}$ if $i \sim j$, and 0 otherwise.
- This scaling ensures that the resulting Ising model is non trivial, for all values of d_n.
- By our scaling, each row sum of A_n equals 1.

Upper bound for M_n

- Thus Perron Frobenius theorem gives $\lambda_\infty(A_n) = 1$.

Upper bound for M_n

- Thus Perron Frobenius theorem gives $\lambda_\infty(A_n) = 1$.
- This gives

$$M_n \leq \sup_{\mathbf{m}\in[-1,1]^n} \left\{ \frac{\beta}{2}\mathbf{m}'\mathbf{m} + \sum_{i=1}^{n} I(m_i) \right\}$$

Upper bound for M_n

- Thus Perron Frobenius theorem gives $\lambda_\infty(A_n) = 1$.
- This gives

$$M_n \leq \sup_{\mathbf{m}\in[-1,1]^n} \left\{ \frac{\beta}{2}\mathbf{m}'\mathbf{m} + \sum_{i=1}^n I(m_i) \right\}$$

$$= \sup_{\mathbf{m}\in[-1,1]^n} \left\{ \frac{\beta}{2}\sum_{i=1}^n m_i^2 + I(m_i) \right\}$$

Upper bound for M_n

- Thus Perron Frobenius theorem gives $\lambda_\infty(A_n) = 1$.
- This gives

$$\begin{aligned} M_n &\leq \sup_{\mathbf{m}\in[-1,1]^n} \left\{ \frac{\beta}{2}\mathbf{m}'\mathbf{m} + \sum_{i=1}^n I(m_i) \right\} \\ &= \sup_{\mathbf{m}\in[-1,1]^n} \left\{ \frac{\beta}{2}\sum_{i=1}^n m_i^2 + I(m_i) \right\} \\ &= \sum_{i=1}^n \sup_{m_i\in[-1,1]} \left\{ \frac{\beta}{2} m_i^2 + I(m_i) \right\} \end{aligned}$$

Upper bound for M_n

- Thus Perron Frobenius theorem gives $\lambda_\infty(A_n) = 1$.
- This gives

$$\begin{aligned} M_n &\leq \sup_{\mathbf{m}\in[-1,1]^n} \left\{ \frac{\beta}{2}\mathbf{m}'\mathbf{m} + \sum_{i=1}^n I(m_i) \right\} \\ &= \sup_{\mathbf{m}\in[-1,1]^n} \left\{ \frac{\beta}{2}\sum_{i=1}^n m_i^2 + I(m_i) \right\} \\ &= \sum_{i=1}^n \sup_{m_i\in[-1,1]} \left\{ \frac{\beta}{2} m_i^2 + I(m_i) \right\} \\ &= n \sup_{m\in[-1,1]} \left\{ \frac{\beta}{2} m^2 + I(m) \right\}. \end{aligned}$$

Lower bound for M_n

- Also, by restricting the sup over $\mathbf{m} \in [-1,1]^n$ such that $m_i = m$ is same for all i, we have

$$M_n \geq \sup_{m \in [-1,1]} \left\{ \frac{\beta}{2} m^2 \mathbf{1}' A_n \mathbf{1} + nI(m) \right\}$$

Lower bound for M_n

- Also, by restricting the sup over $\mathbf{m} \in [-1,1]^n$ such that $m_i = m$ is same for all i, we have

$$\begin{aligned} M_n \geq & \sup_{m \in [-1,1]} \left\{ \frac{\beta}{2} m^2 \mathbf{1}' A_n \mathbf{1} + nI(m) \right\} \\ = & n \sup_{m \in [-1,1]} \left\{ \frac{\beta}{2} m^2 + I(m) \right\}. \end{aligned}$$

Lower bound for M_n

- Also, by restricting the sup over $\mathbf{m} \in [-1,1]^n$ such that $m_i = m$ is same for all i, we have

$$\begin{aligned} M_n \geq & \sup_{m \in [-1,1]} \left\{ \frac{\beta}{2} m^2 \mathbf{1}' A_n \mathbf{1} + nI(m) \right\} \\ = & n \sup_{m \in [-1,1]} \left\{ \frac{\beta}{2} m^2 + I(m) \right\}. \end{aligned}$$

- In the last line we use $A_n \mathbf{1} = \mathbf{1}$.

Lower bound for M_n

- Also, by restricting the sup over $\mathbf{m} \in [-1,1]^n$ such that $m_i = m$ is same for all i, we have

$$\begin{aligned} M_n \geq & \sup_{m \in [-1,1]} \left\{ \frac{\beta}{2} m^2 \mathbf{1}' A_n \mathbf{1} + nI(m) \right\} \\ = & n \sup_{m \in [-1,1]} \left\{ \frac{\beta}{2} m^2 + I(m) \right\}. \end{aligned}$$

- In the last line we use $A_n \mathbf{1} = \mathbf{1}$.

- Thus combining we have

$$M_n = n \sup_{m \in [-1,1]} \left\{ \frac{\beta}{2} m^2 + I(m) \right\}.$$

Solving the optimization

- Consider the optimization of the function $m \mapsto \frac{\beta}{2}m^2 + I(m)$ on $[-1, 1]$.

Solving the optimization

- Consider the optimization of the function $m \mapsto \frac{\beta}{2} m^2 + I(m)$ on $[-1, 1]$.
- Differentiating with respect to m gives the equation $m = \tanh(\beta m)$.

Solving the optimization

- Consider the optimization of the function $m \mapsto \frac{\beta}{2} m^2 + I(m)$ on $[-1, 1]$.
- Differentiating with respect to m gives the equation $m = \tanh(\beta m)$.
- If $\beta \leq 1$ then this equation has a unique solution $m = 0$.

Solving the optimization

- Consider the optimization of the function $m \mapsto \frac{\beta}{2} m^2 + I(m)$ on $[-1, 1]$.
- Differentiating with respect to m gives the equation $m = \tanh(\beta m)$.
- If $\beta \leq 1$ then this equation has a unique solution $m = 0$.
- If $\beta > 1$ then this equation has three solutions, $\{0, m_\beta, -m_\beta\}$.

Solving the optimization

- Consider the optimization of the function $m \mapsto \frac{\beta}{2}m^2 + I(m)$ on $[-1, 1]$.
- Differentiating with respect to m gives the equation $m = \tanh(\beta m)$.
- If $\beta \leq 1$ then this equation has a unique solution $m = 0$.
- If $\beta > 1$ then this equation has three solutions, $\{0, m_\beta, -m_\beta\}$.
- Of these 0 is a local minimizer, and $\pm m_\beta$ are glolal maximizers.

So?

- For any sequence of d_n regular graphs, we have

$$\liminf_{n\to\infty} \frac{1}{n} F_n \geq \sup_{m\in[-1,1]} \left\{ \frac{\beta}{2} m^2 + I(m) \right\}.$$

So?

- For any sequence of d_n regular graphs, we have

$$\liminf_{n\to\infty} \frac{1}{n} F_n \geq \sup_{m\in[-1,1]} \left\{ \frac{\beta}{2} m^2 + I(m) \right\}.$$

- If the graph happens to be a complete graph $(d_n = n-1)$, then using large deviation arguments for i.i.d. random variables one can show the limit is tight.

So?

- For any sequence of d_n regular graphs, we have

$$\liminf_{n\to\infty} \frac{1}{n} F_n \geq \sup_{m\in[-1,1]} \left\{ \frac{\beta}{2} m^2 + I(m) \right\}.$$

- If the graph happens to be a complete graph $(d_n = n-1)$, then using large deviation arguments for i.i.d. random variables one can show the limit is tight.

- It is also known that equality does not hold in the above limit if the graph happens to be a random d_n regular graph with $d_n = d$ fixed.

So?

- For any sequence of d_n regular graphs, we have

$$\liminf_{n\to\infty} \frac{1}{n} F_n \geq \sup_{m\in[-1,1]} \left\{ \frac{\beta}{2} m^2 + I(m) \right\}.$$

- If the graph happens to be a complete graph $(d_n = n-1)$, then using large deviation arguments for i.i.d. random variables one can show the limit is tight.

- It is also known that equality does not hold in the above limit if the graph happens to be a random d_n regular graph with $d_n = d$ fixed.

- This raises the natural question: "For which sequence of regular graphs is the above limit tight?"

Regular graphs continued

- In a regular graph the scaled adjacency matrix has maximum eigenvalue 1, with eigenvector $\frac{1}{\sqrt{n}}\mathbf{1}$.

Regular graphs continued

- In a regular graph the scaled adjacency matrix has maximum eigenvalue 1, with eigenvector $\frac{1}{\sqrt{n}}\mathbf{1}$.
- Assume that all other eigenvalues are $o_n(1)$.

Regular graphs continued

- In a regular graph the scaled adjacency matrix has maximum eigenvalue 1, with eigenvector $\frac{1}{\sqrt{n}}\mathbf{1}$.
- Assume that all other eigenvalues are $o_n(1)$.
- Thus we have $\lambda_\infty\left(A_n - \frac{1}{n}\mathbf{1}\mathbf{1}'\right) = o_n(1)$.

Regular graphs continued

- In a regular graph the scaled adjacency matrix has maximum eigenvalue 1, with eigenvector $\frac{1}{\sqrt{n}}\mathbf{1}$.
- Assume that all other eigenvalues are $o_n(1)$.
- Thus we have $\lambda_\infty\left(A_n - \frac{1}{n}\mathbf{1}\mathbf{1}'\right) = o_n(1)$.
- For any $\mathbf{x} \in \{-1,1\}^n$, this gives

$$\mathbf{x}'\left(A_n - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{x} \approx o_n(1)\mathbf{x}'\mathbf{x} = o(n)$$

Regular graphs continued

- In a regular graph the scaled adjacency matrix has maximum eigenvalue 1, with eigenvector $\frac{1}{\sqrt{n}}\mathbf{1}$.
- Assume that all other eigenvalues are $o_n(1)$.
- Thus we have $\lambda_\infty\left(A_n - \frac{1}{n}\mathbf{1}\mathbf{1}'\right) = o_n(1)$.
- For any $\mathbf{x} \in \{-1,1\}^n$, this gives

$$\mathbf{x}'\left(A_n - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{x} \approx o_n(1)\mathbf{x}'\mathbf{x} = o(n)$$

$$\Rightarrow \mathbf{x}' A_n \mathbf{x} = \frac{1}{n}(\sum_{i=1}^{n} x_i)^2 + o(n)$$

Regular graphs continued

- In a regular graph the scaled adjacency matrix has maximum eigenvalue 1, with eigenvector $\frac{1}{\sqrt{n}}\mathbf{1}$.
- Assume that all other eigenvalues are $o_n(1)$.
- Thus we have $\lambda_\infty\left(A_n - \frac{1}{n}\mathbf{1}\mathbf{1}'\right) = o_n(1)$.
- For any $\mathbf{x} \in \{-1,1\}^n$, this gives

$$\mathbf{x}'\left(A_n - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{x} \approx o_n(1)\mathbf{x}'\mathbf{x} = o(n)$$

$$\Rightarrow \mathbf{x}'A_n\mathbf{x} = \frac{1}{n}(\sum_{i=1}^{n} x_i)^2 + o(n)$$

$$= \frac{1}{n}[n + \sum_{i \neq j} x_i x_j] + o(n)$$

Regular graphs continued

- Summing over $\mathbf{x} \in \{-1, 1\}^n$ we have

$$\sum_{\mathbf{x}\in\{-1,1\}^n} e^{\frac{\beta}{2}\mathbf{x}' A_n \mathbf{x}} \approx e^{o(n)} \sum_{\mathbf{x}\in\{-1,1\}^n} e^{\frac{\beta}{2n}\sum_{i\neq j} x_i x_j}.$$

Regular graphs continued

- Summing over $\mathbf{x} \in \{-1, 1\}^n$ we have

$$\sum_{\mathbf{x}\in\{-1,1\}^n} e^{\frac{\beta}{2}\mathbf{x}' A_n \mathbf{x}} \approx e^{o(n)} \sum_{\mathbf{x}\in\{-1,1\}^n} e^{\frac{\beta}{2n}\sum_{i\neq j} x_i x_j}.$$

- Upto a $o(n)$ term, the RHS above is exactly the partition function of the Curie Weiss model.

Regular graphs continued

- Summing over $\mathbf{x} \in \{-1,1\}^n$ we have

$$\sum_{\mathbf{x}\in\{-1,1\}^n} e^{\frac{\beta}{2}\mathbf{x}' A_n \mathbf{x}} \approx e^{o(n)} \sum_{\mathbf{x}\in\{-1,1\}^n} e^{\frac{\beta}{2n}\sum_{i\neq j} x_i x_j}.$$

- Upto a $o(n)$ term, the RHS above is exactly the partition function of the Curie Weiss model.

- Thus using the Curie Weiss solution we readily get

$$\lim_{n\to\infty} \frac{F_n}{n} = \sup_{m\in[-1,1]} \left\{ \frac{\beta}{2} m^2 + I(m) \right\}.$$

Regular graphs continued

- Summing over $\mathbf{x} \in \{-1,1\}^n$ we have

$$\sum_{\mathbf{x}\in\{-1,1\}^n} e^{\frac{\beta}{2}\mathbf{x}' A_n \mathbf{x}} \approx e^{o(n)} \sum_{\mathbf{x}\in\{-1,1\}^n} e^{\frac{\beta}{2n}\sum_{i\neq j} x_i x_j}.$$

- Upto a $o(n)$ term, the RHS above is exactly the partition function of the Curie Weiss model.

- Thus using the Curie Weiss solution we readily get

$$\lim_{n\to\infty} \frac{F_n}{n} = \sup_{m\in[-1,1]} \Big\{ \frac{\beta}{2} m^2 + I(m) \Big\}.$$

- The above heuristic is rigorous, and so we get the same asymptotics for any regular graph, as soon as the matrix A_n has only one dominant eigenvalue, i.e. there is a spectral gap.

Example: Erdős-Renyi graphs

- Suppose G_n is the adjacency matrix of an Erdős-Renyi graph on n vertices with parameter p_n.

Example: Erdős-Renyi graphs

- Suppose G_n is the adjacency matrix of an Erdős-Renyi graph on n vertices with parameter p_n.
- In this case we choose A_n to be G_n scaled by np_n, which is the average degree.

Example: Erdős-Renyi graphs

- Suppose G_n is the adjacency matrix of an Erdős-Renyi graph on n vertices with parameter p_n.
- In this case we choose A_n to be G_n scaled by np_n, which is the average degree.
- If $p_n \gg \frac{\log n}{n}$ then $\lambda_\infty(G_n) \approx np_n$ (Krivelevich-Sudakov '03), and $\lambda_\infty\left(G_n - p_n \mathbf{1}\mathbf{1}'\right) = O_p(\sqrt{np_n})$(Feige Ofek '05).

Example: Erdős-Renyi graphs

- Suppose G_n is the adjacency matrix of an Erdős-Renyi graph on n vertices with parameter p_n.
- In this case we choose A_n to be G_n scaled by np_n, which is the average degree.
- If $p_n \gg \frac{\log n}{n}$ then $\lambda_\infty(G_n) \approx np_n$ (Krivelevich-Sudakov '03), and $\lambda_\infty\left(G_n - p_n \mathbf{1}\mathbf{1}'\right) = O_p(\sqrt{np_n})$(Feige Ofek '05).
- Thus $\lambda_\infty\left(A_n - \frac{1}{n}\mathbf{1}\mathbf{1}'\right) = O_p\left(\frac{1}{\sqrt{np_n}}\right) = o_n(1)$, as $np_n \gg \log n$.

Example: Erdős-Renyi graphs

- Thus it follows that

$$\lim_{n\to\infty} \frac{F_n}{n} = \sup_{m\in[-1,1]} \left\{ \frac{\beta}{2} m^2 + I(m) \right\}.$$

Example: Erdős-Renyi graphs

- Thus it follows that

$$\lim_{n\to\infty} \frac{F_n}{n} = \sup_{m\in[-1,1]} \left\{ \frac{\beta}{2} m^2 + I(m) \right\}.$$

- If $p_n \ll \frac{\log n}{n}$ then estimates for $\lambda_1(G_n)$ and $\lambda_\infty(G_n - p\mathbf{1}\mathbf{1}')$ are not as nice.

Example: Erdős-Renyi graphs

- Thus it follows that

$$\lim_{n\to\infty} \frac{F_n}{n} = \sup_{m\in[-1,1]} \left\{ \frac{\beta}{2} m^2 + I(m) \right\}.$$

- If $p_n \ll \frac{\log n}{n}$ then estimates for $\lambda_1(G_n)$ and $\lambda_\infty(G_n - p\mathbf{1}\mathbf{1}')$ are not as nice.

- Also we know mean field prediction does not hold when np_n is constant.

Example: Erdős-Renyi graphs

- Thus it follows that

$$\lim_{n\to\infty} \frac{F_n}{n} = \sup_{m\in[-1,1]} \left\{ \frac{\beta}{2} m^2 + I(m) \right\}.$$

- If $p_n \ll \frac{\log n}{n}$ then estimates for $\lambda_1(G_n)$ and $\lambda_\infty(G_n - p\mathbf{1}\mathbf{1}')$ are not as nice.

- Also we know mean field prediction does not hold when np_n is constant.

- Thus a natural question is where does the mean field approximation break down.

Example: Random d_n regular graphs

- Suppose G_n is the adjacency matrix of a random d_n regular graph, and A_n is G_n scaled by d_n.

Example: Random d_n regular graphs

- Suppose G_n is the adjacency matrix of a random d_n regular graph, and A_n is G_n scaled by d_n.
- In this case $\lambda_1(G_n) = d_n$, and $\lambda_2(G_n) = o(d_n)$ as soon as $d_n \geq (\log n)^\gamma$ for some γ (Chung-Lu-Vu '03, Coja-Oghlan-Lanka '09).

Example: Random d_n regular graphs

- Suppose G_n is the adjacency matrix of a random d_n regular graph, and A_n is G_n scaled by d_n.
- In this case $\lambda_1(G_n) = d_n$, and $\lambda_2(G_n) = o(d_n)$ as soon as $d_n \geq (\log n)^\gamma$ for some γ (Chung-Lu-Vu '03, Coja-Oghlan-Lanka '09).
- Thus a similar argument goes through, proving that mean field prediction is asymptotically correct if $d_n \geq (\log n)^\gamma$.

Example: Random d_n regular graphs

- Suppose G_n is the adjacency matrix of a random d_n regular graph, and A_n is G_n scaled by d_n.
- In this case $\lambda_1(G_n) = d_n$, and $\lambda_2(G_n) = o(d_n)$ as soon as $d_n \geq (\log n)^\gamma$ for some γ (Chung-Lu-Vu '03, Coja-Oghlan-Lanka '09).
- Thus a similar argument goes through, proving that mean field prediction is asymptotically correct if $d_n \geq (\log n)^\gamma$.
- Again there is a question of how far one can stretch this result.

Example: Random d_n regular graphs

- Suppose G_n is the adjacency matrix of a random d_n regular graph, and A_n is G_n scaled by d_n.
- In this case $\lambda_1(G_n) = d_n$, and $\lambda_2(G_n) = o(d_n)$ as soon as $d_n \geq (\log n)^\gamma$ for some γ (Chung-Lu-Vu '03, Coja-Oghlan-Lanka '09).
- Thus a similar argument goes through, proving that mean field prediction is asymptotically correct if $d_n \geq (\log n)^\gamma$.
- Again there is a question of how far one can stretch this result. It is known that the mean field prediction is not correct if $d_n = d$ is fixed.

What is the underlying picture?

- If the random graph is dense enough, then the mean field prediction is expected to hold via dominant eigenvalue argument.

What is the underlying picture?

- If the random graph is dense enough, then the mean field prediction is expected to hold via dominant eigenvalue argument.

- How dense is dense enough depends on the underlying model.

What is the underlying picture?

- If the random graph is dense enough, then the mean field prediction is expected to hold via dominant eigenvalue argument.
- How dense is dense enough depends on the underlying model.
- None of the results mentioned above applies for a specific sequence of non random graphs.

What is the underlying picture?

- If the random graph is dense enough, then the mean field prediction is expected to hold via dominant eigenvalue argument.
- How dense is dense enough depends on the underlying model.
- None of the results mentioned above applies for a specific sequence of non random graphs.
- However, the mean field prediction has this very nice form for any regular graph.

Example: The Hypercube

- Suppose G_n is the adjacency matrix of a d dimensional hypercube $\{0,1\}^d$.

Example: The Hypercube

- Suppose G_n is the adjacency matrix of a d dimensional hypercube $\{0,1\}^d$.
- In this case the number of vertices is $n = 2^d$.

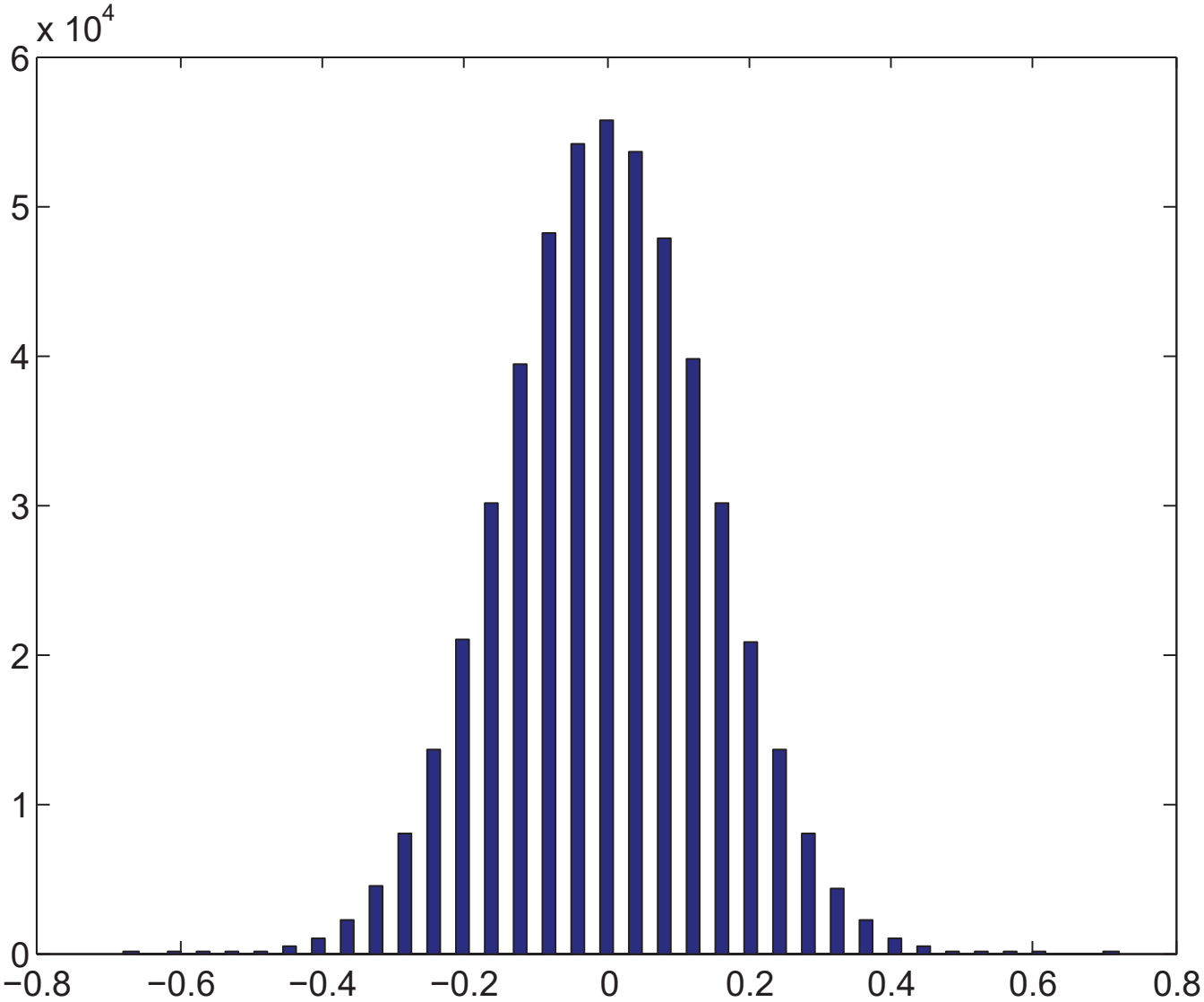
Example: The Hypercube

- Suppose G_n is the adjacency matrix of a d dimensional hypercube $\{0,1\}^d$.
- In this case the number of vertices is $n = 2^d$.
- G_n has eigenvalues $d - 2i$ with multiplicity $\binom{d}{i}$, for $0 \leq i \leq d$.

Example: The Hypercube

- Suppose G_n is the adjacency matrix of a d dimensional hypercube $\{0,1\}^d$.
- In this case the number of vertices is $n = 2^d$.
- G_n has eigenvalues $d-2i$ with multiplicity $\binom{d}{i}$, for $0 \leq i \leq d$.
- Thus the eigenvalues of G_n range between $[-d, d]$, and there is no dominant eigenvalue.

Example: The Hypercube ($d = 50$)



Example: The hypercube

- In this case $d = d_n = \log_2 n$, which is somewhere between sparse and dense graphs.

Example: The hypercube

- In this case $d = d_n = \log_2 n$, which is somewhere between sparse and dense graphs.
- There is essentially no spectral gap, and hence dominant eigenvalue argument fails.

Example: The hypercube

- In this case $d = d_n = \log_2 n$, which is somewhere between sparse and dense graphs.
- There is essentially no spectral gap, and hence dominant eigenvalue argument fails.
- Thus it is not clear even at a heuristic level whether the mean field prediction should be correct.

A more general question

- More generally, for a general sequence of matrices A_n (not necessarily an adjacency matrix), we pose the question of when is the mean field prediction tight.

A more general question

- More generally, for a general sequence of matrices A_n (not necessarily an adjacency matrix), we pose the question of when is the mean field prediction tight.

- Thus we look for sufficient condition on the sequence of matrices A_n such that

$$\lim_{n\to\infty} \frac{1}{n}(F_n - M_n) = 0.$$

A more general question

- More generally, for a general sequence of matrices A_n (not necessarily an adjacency matrix), we pose the question of when is the mean field prediction tight.

- Thus we look for sufficient condition on the sequence of matrices A_n such that

$$\lim_{n\to\infty} \frac{1}{n}(F_n - M_n) = 0.$$

- If this holds, then we are able to reduce the asymptotics of F_n to the asymptotics of M_n.

A more general question

- More generally, for a general sequence of matrices A_n (not necessarily an adjacency matrix), we pose the question of when is the mean field prediction tight.

- Thus we look for sufficient condition on the sequence of matrices A_n such that

$$\lim_{n\to\infty} \frac{1}{n}(F_n - M_n) = 0.$$

- If this holds, then we are able to reduce the asymptotics of F_n to the asymptotics of M_n.

- A follow up question is how to solve the optimization problem in M_n.

A more general question

- More generally, for a general sequence of matrices A_n (not necessarily an adjacency matrix), we pose the question of when is the mean field prediction tight.

- Thus we look for sufficient condition on the sequence of matrices A_n such that

$$\lim_{n\to\infty} \frac{1}{n}(F_n - M_n) = 0.$$

- If this holds, then we are able to reduce the asymptotics of F_n to the asymptotics of M_n.

- A follow up question is how to solve the optimization problem in M_n. As we checked, the optimization problem is easy for regular graphs.

The proposed solution

- We claim that a sufficient condition for the mean field prediction to hold is

$$\sum_{i,j=1}^{n} A_n(i,j)^2 = \mathrm{tr}(A_n^2) = \sum_{i=1}^{n} \lambda_i^2 = o(n).$$

The proposed solution

- We claim that a sufficient condition for the mean field prediction to hold is

$$\sum_{i,j=1}^{n} A_n(i,j)^2 = \mathrm{tr}(A_n^2) = \sum_{i=1}^{n} \lambda_i^2 = o(n).$$

- This basically says that the empirical eigenvalue distribution converges to 0 in L^2.

The proposed solution

- We claim that a sufficient condition for the mean field prediction to hold is

$$\sum_{i,j=1}^{n} A_n(i,j)^2 = \mathrm{tr}(A_n^2) = \sum_{i=1}^{n} \lambda_i^2 = o(n).$$

- This basically says that the empirical eigenvalue distribution converges to 0 in L^2.

- One way to think about this claim is that in this case most of the eigenvalues are $o(1)$.

The proposed solution

- We claim that a sufficient condition for the mean field prediction to hold is

$$\sum_{i,j=1}^{n} A_n(i,j)^2 = \mathrm{tr}(A_n^2) = \sum_{i=1}^{n} \lambda_i^2 = o(n).$$

- This basically says that the empirical eigenvalue distribution converges to 0 in L^2.

- One way to think about this claim is that in this case most of the eigenvalues are $o(1)$.

- Thus the number of dominant eigenvalues is $o(n)$.

Example: Ising models on graphs

- In this case A_n is the adjacency matrix of G_n scaled by the average degree $\frac{2E(G_n)}{n}$.

Example: Ising models on graphs

- In this case A_n is the adjacency matrix of G_n scaled by the average degree $\frac{2E(G_n)}{n}$.

- Thus we have

$$\sum_{i,j=1}^{n} A_n^2(i,j) = \frac{n^2}{4E(G_n)^2} \sum_{i,j=1}^{n} 1\{i \sim j\} = \frac{n^2}{2E(G_n)}.$$

Example: Ising models on graphs

- In this case A_n is the adjacency matrix of G_n scaled by the average degree $\frac{2E(G_n)}{n}$.
- Thus we have

$$\sum_{i,j=1}^{n} A_n^2(i,j) = \frac{n^2}{4E(G_n)^2} \sum_{i,j=1}^{n} 1\{i \sim j\} = \frac{n^2}{2E(G_n)}.$$

- The RHS is $o(n)$ iff $E(G_n) \gg n$.

Example: Ising models on graphs

- In this case A_n is the adjacency matrix of G_n scaled by the average degree $\frac{2E(G_n)}{n}$.
- Thus we have

$$\sum_{i,j=1}^{n} A_n^2(i,j) = \frac{n^2}{4E(G_n)^2} \sum_{i,j=1}^{n} 1\{i \sim j\} = \frac{n^2}{2E(G_n)}.$$

- The RHS is $o(n)$ iff $E(G_n) \gg n$.
- Thus mean field prediction holds as soon as the graph has super linear edges.

Example: Erdős-Renyi graphs

- In this case $2E(G_n) \approx n^2 p_n$.

Example: Erdős-Renyi graphs

- In this case $2E(G_n) \approx n^2 p_n$.
- This is super linear as soon as $np_n \to \infty$.

Example: Erdős-Renyi graphs

- In this case $2E(G_n) \approx n^2 p_n$.
- This is super linear as soon as $np_n \to \infty$.
- Also if $np_n = c$ is fixed, then mean field prediction does not hold.

Example: Erdős-Renyi graphs

- In this case $2E(G_n) \approx n^2 p_n$.
- This is super linear as soon as $np_n \to \infty$.
- Also if $np_n = c$ is fixed, then mean field prediction does not hold.
- This gives a complete picture for Erdős-Renyi graphs.

Example: Regular graphs

- In this case $2E(G_n) = nd_n$.

Example: Regular graphs

- In this case $2E(G_n) = nd_n$.
- This is super linear, as soon as $d_n \to \infty$.

Example: Regular graphs

- In this case $2E(G_n) = nd_n$.
- This is super linear, as soon as $d_n \to \infty$.
- Thus mean field prediction holds for any sequence of d_n regular graphs, as soon as $d_n \to \infty$.

Example: Regular graphs

- In this case $2E(G_n) = nd_n$.
- This is super linear, as soon as $d_n \to \infty$.
- Thus mean field prediction holds for any sequence of d_n regular graphs, as soon as $d_n \to \infty$. In particular this covers the hypercube, as $d_n = \log_2 n$.

Example: Regular graphs

- In this case $2E(G_n) = nd_n$.
- This is super linear, as soon as $d_n \to \infty$.
- Thus mean field prediction holds for any sequence of d_n regular graphs, as soon as $d_n \to \infty$. In particular this covers the hypercube, as $d_n = \log_2 n$.
- For $d_n = d$ fixed mean field prediction does not hold.

Sanity check: SK Model

- In this case $A_n(i,j) = \frac{1}{\sqrt{n}} Z(i,j)$, where $\{Z(i,j), 1 \le i < j \le n\} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$.

Sanity check: SK Model

- In this case $A_n(i,j) = \frac{1}{\sqrt{n}} Z(i,j)$, where $\{Z(i,j), 1 \le i < j \le n\} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$.

- This gives

$$\frac{1}{n} \sum_{i,j=1}^{n} A_n(i,j)^2 = \frac{1}{n^2} \sum_{i,j=1}^{n} Z(i,j)^2 \xrightarrow{p} 1.$$

Sanity check: SK Model

- In this case $A_n(i,j) = \frac{1}{\sqrt{n}} Z(i,j)$, where $\{Z(i,j), 1 \le i < j \le n\} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$.
- This gives

$$\frac{1}{n} \sum_{i,j=1}^{n} A_n(i,j)^2 = \frac{1}{n^2} \sum_{i,j=1}^{n} Z(i,j)^2 \xrightarrow{p} 1.$$

- Thus we have $\sum_{i,j=1}^{n} A_n(i,j)^2 = \Theta_p(n)$, and so mean field condition does not hold.

Sanity check: SK Model

- In this case $A_n(i,j) = \frac{1}{\sqrt{n}} Z(i,j)$, where $\{Z(i,j), 1 \leq i < j \leq n\} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$.
- This gives

$$\frac{1}{n} \sum_{i,j=1}^{n} A_n(i,j)^2 = \frac{1}{n^2} \sum_{i,j=1}^{n} Z(i,j)^2 \xrightarrow{p} 1.$$

- Thus we have $\sum_{i,j=1}^{n} A_n(i,j)^2 = \Theta_p(n)$, and so mean field condition does not hold.
- This is what should be the case, as the mean field prediction does not hold for the SK Model.

Comments about our results

- Our results allow for the presence of an external magnetic field term of the form $B \sum_{i=1}^{n} x_i$ in the Ising model.

Comments about our results

- Our results allow for the presence of an external magnetic field term of the form $B \sum_{i=1}^{n} x_i$ in the Ising model.

- Our results apply for Potts model as well, where we have q states instead of 2.

Comments about our results

- Our results allow for the presence of an external magnetic field term of the form $B\sum_{i=1}^{n} x_i$ in the Ising model.
- Our results apply for Potts model as well, where we have q states instead of 2.
- There is a technical condition required on the matrix A_n, which is

$$\sup_{\mathbf{x}\in[-1,1]^n} \sum_{i=1}^{n} \left| \sum_{j=1}^{n} x_i A_n(i,j) \right| = O(n).$$

Comments about our results

- Our results allow for the presence of an external magnetic field term of the form $B\sum_{i=1}^{n} x_i$ in the Ising model.
- Our results apply for Potts model as well, where we have q states instead of 2.
- There is a technical condition required on the matrix A_n, which is

$$\sup_{\mathbf{x}\in[-1,1]^n} \sum_{i=1}^{n} \left| \sum_{j=1}^{n} x_i A_n(i,j) \right| = O(n).$$

- We don't know of a single model in the literature which violates this assumption, be it mean field or otherwise

Comments about our results

- Our results allow for the presence of an external magnetic field term of the form $B\sum_{i=1}^{n} x_i$ in the Ising model.
- Our results apply for Potts model as well, where we have q states instead of 2.
- There is a technical condition required on the matrix A_n, which is

$$\sup_{\mathbf{x}\in[-1,1]^n} \sum_{i=1}^{n} \left| \sum_{j=1}^{n} x_i A_n(i,j) \right| = O(n).$$

- We don't know of a single model in the literature which violates this assumption, be it mean field or otherwise (Ising model on all graphs, SK Model, Hopfield Model).

What about the optimization problem?

- Our general sufficient condition gives a wide class of graphs for which the mean field prediction holds.

What about the optimization problem?

- Our general sufficient condition gives a wide class of graphs for which the mean field prediction holds.

- This means that for Ising models on such graphs we have $\lim_{n\to\infty} \frac{1}{n}(F_n - M_n) = 0$.

What about the optimization problem?

- Our general sufficient condition gives a wide class of graphs for which the mean field prediction holds.
- This means that for Ising models on such graphs we have $\lim_{n\to\infty} \frac{1}{n}(F_n - M_n) = 0$.
- Recall that

$$M_n = \sup_{\mathbf{m}\in[-1,1]^n} \left\{ \frac{1}{2}\mathbf{m}' A_n \mathbf{m} + \sum_{i=1}^{n} I(m_i) \right\}.$$

What about the optimization problem?

- Our general sufficient condition gives a wide class of graphs for which the mean field prediction holds.
- This means that for Ising models on such graphs we have $\lim_{n\to\infty} \frac{1}{n}(F_n - M_n) = 0$.
- Recall that

$$M_n = \sup_{\mathbf{m}\in[-1,1]^n} \left\{ \frac{1}{2}\mathbf{m}' A_n \mathbf{m} + \sum_{i=1}^{n} I(m_i) \right\}.$$

- Solving this for regular graphs was easy.

What about the optimization problem?

- Our general sufficient condition gives a wide class of graphs for which the mean field prediction holds.
- This means that for Ising models on such graphs we have $\lim_{n\to\infty} \frac{1}{n}(F_n - M_n) = 0$.
- Recall that

$$M_n = \sup_{\mathbf{m}\in[-1,1]^n} \left\{ \frac{1}{2}\mathbf{m}' A_n \mathbf{m} + \sum_{i=1}^{n} I(m_i) \right\}.$$

- Solving this for regular graphs was easy. What can we say about other graphs?

Approximately regular graphs

- Suppose all the degrees d_i are approximately close to $\bar{d}$, with $\bar{d} \to \infty$.

Approximately regular graphs

- Suppose all the degrees d_i are approximately close to $\bar{d}$, with $\bar{d} \to \infty$.
- Then one expects the same asymptotics for F_n as in the regular graph case.

Approximately regular graphs

- Suppose all the degrees d_i are approximately close to $\bar{d}$, with $\bar{d} \to \infty$.
- Then one expects the same asymptotics for F_n as in the regular graph case.
- We show this is the case under the assumption that the empirical distribution $\sum_{i=1}^{n} \frac{1}{n} \delta_{\frac{d_i}{\bar{d}}} \xrightarrow{w} \delta_1$.

Approximately regular graphs

- Suppose all the degrees d_i are approximately close to $\bar{d}$, with $\bar{d} \to \infty$.
- Then one expects the same asymptotics for F_n as in the regular graph case.
- We show this is the case under the assumption that the empirical distribution $\sum_{i=1}^{n} \frac{1}{n} \delta_{\frac{d_i}{\bar{d}}} \xrightarrow{w} \delta_1$.
- As a by product of our analysis, we show that $\bar{X}$ has the same large deviation for all approximately regular graphs in the Ferro-magnetic domain $(\beta > 0)$, as soon as the average degree goes to $+\infty$.

Ising model on bi-regular bi-partite graphs

- Let G_n be a sequence of bi-regular bi-partite graphs with bipartition sizes a_n and b_n.

Ising model on bi-regular bi-partite graphs

- Let G_n be a sequence of bi-regular bi-partite graphs with bipartition sizes a_n and b_n.
- Thus we have $a_n + b_n = n$, the total number of vertices.

Ising model on bi-regular bi-partite graphs

- Let G_n be a sequence of bi-regular bi-partite graphs with bipartition sizes a_n and b_n.
- Thus we have $a_n + b_n = n$, the total number of vertices.
- Assume that $\lim_{n\to\infty} \frac{1}{n} a_n = p \in (0, 1)$.

Ising model on bi-regular bi-partite graphs

- Let G_n be a sequence of bi-regular bi-partite graphs with bipartition sizes a_n and b_n.
- Thus we have $a_n + b_n = n$, the total number of vertices.
- Assume that $\lim_{n\to\infty} \frac{1}{n} a_n = p \in (0, 1)$.
- This removes the possibility that one of the partitions dominate.

Ising model on bi-regular bi-partite graphs

- Assume that each bi-partition has constant degree c_n and d_n respectively.

Ising model on bi-regular bi-partite graphs

- Assume that each bi-partition has constant degree c_n and d_n respectively.
- Then we must have $a_n c_n = b_n d_n$, which equals the total number of edges.

Ising model on bi-regular bi-partite graphs

- Assume that each bi-partition has constant degree c_n and d_n respectively.
- Then we must have $a_n c_n = b_n d_n$, which equals the total number of edges.
- Thus for $E(G_n)$ to be super linear, we need $c_n($ or equivalently $d_n)$ to go to $+\infty$.

Ising model on bi-regular bi-partite graphs

- Choose A_n to be the adjacency matrix of G_n scaled by $\frac{1}{c_n+d_n}$.

Ising model on bi-regular bi-partite graphs

- Choose A_n to be the adjacency matrix of G_n scaled by $\frac{1}{c_n+d_n}$.

- The asymptotics of $\frac{F_n}{n}$ is given by

$$\sup_{m_1,m_2\in[-1,1]} \{\beta p(1-p)m_1m_2 + pI(m_1) + (1-p)I(m_2)\}.$$

Ising model on bi-regular bi-partite graphs

- Choose A_n to be the adjacency matrix of G_n scaled by $\frac{1}{c_n+d_n}$.

- The asymptotics of $\frac{F_n}{n}$ is given by

$$\sup_{m_1,m_2\in[-1,1]} \{\beta p(1-p)m_1m_2 + pI(m_1) + (1-p)I(m_2)\}.$$

- Thus again the asymptotics is universal for any sequence of bi-regular bi-partite graphs, as soon as the average degree goes to $+\infty$.

Ising model on bi-regular bi-partite graphs

- Choose A_n to be the adjacency matrix of G_n scaled by $\frac{1}{c_n+d_n}$.

- The asymptotics of $\frac{F_n}{n}$ is given by

$$\sup_{m_1,m_2\in[-1,1]} \{\beta p(1-p)m_1m_2 + pI(m_1) + (1-p)I(m_2)\}.$$

- Thus again the asymptotics is universal for any sequence of bi-regular bi-partite graphs, as soon as the average degree goes to $+\infty$.

- Differentiating with respect to m_1, m_2 we get the equations $m_1 = \tanh(\beta(1-p)m_2),\ m_2 = \tanh(\beta p m_1)$.

Ising model on bi-regular bi-partite graphs

- If $|\beta| \leq \frac{1}{\sqrt{p(1-p)}}$ then there is a unique solution $(m_1, m_2) = (0, 0)$.

Ising model on bi-regular bi-partite graphs

- If $|\beta| \leq \frac{1}{\sqrt{p(1-p)}}$ then there is a unique solution $(m_1, m_2) = (0, 0)$.
- If $\beta > \frac{1}{\sqrt{p(1-p)}}$, then there are two solutions $(x_{\beta,p}, y_{\beta,p})$ and $(-x_{\beta,p}, -y_{\beta,p})$.

Ising model on bi-regular bi-partite graphs

- If $|\beta| \leq \frac{1}{\sqrt{p(1-p)}}$ then there is a unique solution $(m_1, m_2) = (0, 0)$.
- If $\beta > \frac{1}{\sqrt{p(1-p)}}$, then there are two solutions $(x_{\beta,p}, y_{\beta,p})$ and $(-x_{\beta,p}, -y_{\beta,p})$.
- If $\beta < -\frac{1}{\sqrt{p(1-p)}}$, then there are again two solutions $(x_{\beta,p}, -y_{\beta,p})$ and $(-x_{\beta,p}, y_{\beta,p})$.

Difference between K_n and $K_{n/2,n/2}$

- Consider Ising model on two graphs, the complete graph and the complete bi-partite graph.

Difference between K_n and $K_{n/2,n/2}$

- Consider Ising model on two graphs, the complete graph and the complete bi-partite graph.
- If $\beta > 0$ then both the models show same asymtotic behavior.

Difference between K_n and $K_{n/2,n/2}$

- Consider Ising model on two graphs, the complete graph and the complete bi-partite graph.
- If $\beta > 0$ then both the models show same asymtotic behavior. In particular, there is a phase transition in β.

Difference between K_n and $K_{n/2,n/2}$

- Consider Ising model on two graphs, the complete graph and the complete bi-partite graph.
- If $\beta > 0$ then both the models show same asymtotic behavior. In particular, there is a phase transition in β.
- If $\beta < 0$ then the two models behave differently.

Difference between K_n and $K_{n/2,n/2}$

- Consider Ising model on two graphs, the complete graph and the complete bi-partite graph.
- If $\beta > 0$ then both the models show same asymtotic behavior. In particular, there is a phase transition in β.
- If $\beta < 0$ then the two models behave differently.
- Ising model on the complete bi-partite graph shows a phase transition, whereas that on the complete graph does not.

Graphs converging in cut metric

- Given a graph G_n on n vertices, one can define a function W_{G_n} on the unit square as follows:

Graphs converging in cut metric

- Given a graph G_n on n vertices, one can define a function W_{G_n} on the unit square as follows:

- Partition the unit square into n^2 squares each of equal size.

Graphs converging in cut metric

- Given a graph G_n on n vertices, one can define a function W_{G_n} on the unit square as follows:
- Partition the unit square into n^2 squares each of equal size.
- Define W_{G_n} to be equal to 1 on the $(i,j)^{th}$ box if (i,j) is an edge in G_n, and 0 otherwise.

Graphs converging in cut metric

- Given a graph G_n on n vertices, one can define a function W_{G_n} on the unit square as follows:
- Partition the unit square into n^2 squares each of equal size.
- Define W_{G_n} to be equal to 1 on the $(i,j)^{th}$ box if (i,j) is an edge in G_n, and 0 otherwise.
- Thus we have W_{G_n} is a function on the unit square, with $\int_{[0,1]^2} |W_{G_n}(x,y)|dxdy = \frac{2E(G_n)}{n^2}$.

Graphs converging in cut metric

- To proceed in a manner which works across edge densities, scale the function W_{G_n} by $\frac{2E(G_n)}{n^2}$ and call the resulting function by $\widetilde{W}_{G_n}$.

Graphs converging in cut metric

- To proceed in a manner which works across edge densities, scale the function W_{G_n} by $\frac{2E(G_n)}{n^2}$ and call the resulting function by $\widetilde{W}_{G_n}$.

- Thus after scaling we have $\int_{[0,1]^2} |\widetilde{W}_{G_n}(x,y)|dxdy = 1$.

Graphs converging in cut metric

- To proceed in a manner which works across edge densities, scale the function W_{G_n} by $\frac{2E(G_n)}{n^2}$ and call the resulting function by $\widetilde{W}_{G_n}$.
- Thus after scaling we have $\int_{[0,1]^2} |\widetilde{W}_{G_n}(x,y)|dxdy = 1$.
- Consider the space of symmetric measurable functions $W \in L^1[0,1]^2$, to be henceforth referred as graphons.

Graphs converging in cut metric

- To proceed in a manner which works across edge densities, scale the function W_{G_n} by $\frac{2E(G_n)}{n^2}$ and call the resulting function by $\widetilde{W}_{G_n}$.
- Thus after scaling we have $\int_{[0,1]^2} |\widetilde{W}_{G_n}(x,y)|dxdy = 1$.
- Consider the space of symmetric measurable functions $W \in L^1[0,1]^2$, to be henceforth referred as graphons.
- For any graphon W, define the cut norm $||W||_\square = \sup_{S,T \subset [0,1]} |\int_{S \times T} W(x,y)dxdy|$.

Graphs converging in cut metric

- We will say a sequence of graphs G_n converge in cut metric to W, if the functions $\widetilde{W}_{G_n}$ converge in cut norm to W.

Graphs converging in cut metric

- We will say a sequence of graphs G_n converge in cut metric to W, if the functions $\widetilde{W}_{G_n}$ converge in cut norm to W.
- As example, the complete graph converges to $W_1 \equiv 1$.

Graphs converging in cut metric

- We will say a sequence of graphs G_n converge in cut metric to W, if the functions $\widetilde{W}_{G_n}$ converge in cut norm to W.
- As example, the complete graph converges to $W_1 \equiv 1$.
- It follows from (Borgs et. al '14) that if a sequence of graphs G_n converge W in cut metric, then $|E(G_n)| \gg n$.

Graphs converging in cut metric

- We will say a sequence of graphs G_n converge in cut metric to W, if the functions $\widetilde{W}_{G_n}$ converge in cut norm to W.
- As example, the complete graph converges to $W_1 \equiv 1$.
- It follows from (Borgs et. al '14) that if a sequence of graphs G_n converge W in cut metric, then $|E(G_n)| \gg n$.
- Thus the mean field condition holds in this case, and so it suffices to consider the asymptotics of M_n.

Graphs converging in cut metric

- By a simple analysis of M_n, the limiting log partition function is

$$\sup_{m(.)} \Big\{ \int_{[0,1]^2} \frac{\beta}{2} m(x)m(y)W(x,y)dxdy + \int_{[0,1]} I(m(x))dx \Big\}.$$

Graphs converging in cut metric

- By a simple analysis of M_n, the limiting log partition function is

$$\sup_{m(.)} \Big\{ \int_{[0,1]^2} \frac{\beta}{2} m(x)m(y)W(x,y)dxdy + \int_{[0,1]} I(m(x))dx \Big\}.$$

- Here the supremum is over all measurable functions $m : [0,1] \mapsto [-1,1]$.

Graphs converging in cut metric

- By a simple analysis of M_n, the limiting log partition function is

$$\sup_{m(.)} \Big\{ \int_{[0,1]^2} \frac{\beta}{2} m(x)m(y)W(x,y)dxdy + \int_{[0,1]} I(m(x))dx \Big\}.$$

- Here the supremum is over all measurable functions $m : [0,1] \mapsto [-1,1]$.

- This result was already proved in (Borgs et. al '14).

Graphs converging in cut metric

- By a simple analysis of M_n, the limiting log partition function is

$$\sup_{m(.)} \Big\{ \int_{[0,1]^2} \frac{\beta}{2} m(x)m(y)W(x,y)dxdy + \int_{[0,1]} I(m(x))dx \Big\}.$$

- Here the supremum is over all measurable functions $m : [0,1] \mapsto [-1,1]$.

- This result was already proved in (Borgs et. al '14).

- In our paper we re-derive this result as a consequence of mean field prediction, to illustrate the flexibility of this approach.

Graphs converging in cut metric

- By a simple analysis of M_n, the limiting log partition function is

$$\sup_{m(.)} \Big\{ \int_{[0,1]^2} \frac{\beta}{2} m(x)m(y)W(x,y)dxdy + \int_{[0,1]} I(m(x))dx \Big\}.$$

- Here the supremum is over all measurable functions $m : [0,1] \mapsto [-1,1]$.

- This result was already proved in (Borgs et. al '14).

- In our paper we re-derive this result as a consequence of mean field prediction, to illustrate the flexibility of this approach.

- As an immediate application, we obtain the phase transition point for such Ising models, which is $\beta = \frac{1}{\lambda_1(W)} \approx \frac{1}{\lambda_1(A_n)}$.

Proof Strategy

- The main tool for proving the mean field prediction is a large deviation estimate for binary variables (Chatterjee-Dembo '14).

Proof Strategy

- The main tool for proving the mean field prediction is a large deviation estimate for binary variables (Chatterjee-Dembo '14).
- They give a general technique to prove mean field type predictions in exponential families of the form $e^{f(\mathbf{x})}$ on binary random variables $\mathbf{x} \in \{-1, 1\}^n$.

Proof Strategy

- The main tool for proving the mean field prediction is a large deviation estimate for binary variables (Chatterjee-Dembo '14).
- They give a general technique to prove mean field type predictions in exponential families of the form $e^{f(\mathbf{x})}$ on binary random variables $\mathbf{x} \in \{-1, 1\}^n$.
- The idea is that if the gradient of the function $\nabla f(\mathbf{x})$ can be expressed in $o(n)$ bits, then mean field prediction works.

Proof Strategy

- The main tool for proving the mean field prediction is a large deviation estimate for binary variables (Chatterjee-Dembo '14).
- They give a general technique to prove mean field type predictions in exponential families of the form $e^{f(\mathbf{x})}$ on binary random variables $\mathbf{x} \in \{-1, 1\}^n$.
- The idea is that if the gradient of the function $\nabla f(\mathbf{x})$ can be expressed in $o(n)$ bits, then mean field prediction works.
- In our case $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}' A_n \mathbf{x}$ for a symmetric matrix A_n.

E.g. Ising model on Complete graph

- For a specific example take $f(\mathbf{x}) = \frac{1}{n} \sum_{1 \le i < j \le n} x_i x_j$.

E.g. Ising model on Complete graph

- For a specific example take $f(\mathbf{x}) = \frac{1}{n} \sum_{1 \le i < j \le n} x_i x_j$.
- This is the Curie-Weiss Ising model on the complete graph.

E.g. Ising model on Complete graph

- For a specific example take $f(\mathbf{x}) = \frac{1}{n} \sum_{1 \le i < j \le n} x_i x_j$.
- This is the Curie-Weiss Ising model on the complete graph.
- A direct calculation gives

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \frac{1}{n} \sum_{j \neq i} x_j = \bar{x} - \frac{x_i}{n}.$$

E.g. Ising model on Complete graph

- For a specific example take $f(\mathbf{x}) = \frac{1}{n} \sum_{1 \le i < j \le n} x_i x_j$.
- This is the Curie-Weiss Ising model on the complete graph.
- A direct calculation gives

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \frac{1}{n} \sum_{j \neq i} x_j = \bar{x} - \frac{x_i}{n}.$$

- This is approximately $\bar{x}$, and so every component of $\nabla f(\mathbf{x})$ can be expressed by one bit of information $\bar{x}$.

E.g. Ising model on Z^1

- For an example of a model that is not mean field, take $f(\mathbf{x}) = \sum_{i=1}^{n-1} x_i x_{i+1}$.

E.g. Ising model on Z^1

- For an example of a model that is not mean field, take $f(\mathbf{x}) = \sum_{i=1}^{n-1} x_i x_{i+1}$.
- This is the Ising model on the one dimensional integer lattice.

E.g. Ising model on Z^1

- For an example of a model that is not mean field, take $f(\mathbf{x}) = \sum_{i=1}^{n-1} x_i x_{i+1}$.
- This is the Ising model on the one dimensional integer lattice.
- In this case we have

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = x_{i-1} + x_{i+1}$$

for $2 \leq i \leq n-1$.

E.g. Ising model on Z^1

- For an example of a model that is not mean field, take $f(\mathbf{x}) = \sum_{i=1}^{n-1} x_i x_{i+1}$.
- This is the Ising model on the one dimensional integer lattice.
- In this case we have

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = x_{i-1} + x_{i+1}$$

for $2 \leq i \leq n-1$.

- This vector is no longer expressable in terms of $o(n)$ many quantities.

E.g. Ising model on Z^1

- For an example of a model that is not mean field, take $f(\mathbf{x}) = \sum_{i=1}^{n-1} x_i x_{i+1}$.
- This is the Ising model on the one dimensional integer lattice.
- In this case we have

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = x_{i-1} + x_{i+1}$$

  for $2 \leq i \leq n-1$.
- This vector is no longer expressable in terms of $o(n)$ many quantities.
- Thus Ising model on complete graph is mean field, whereas Ising model on Z^1 is not.

Extending to general matrices

- In general if $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}' A_n \mathbf{x}$ then $\nabla f(\mathbf{x}) = A_n \mathbf{x}$.

Extending to general matrices

- In general if $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}' A_n \mathbf{x}$ then $\nabla f(\mathbf{x}) = A_n \mathbf{x}$.
- Assume A_n has spectral decomposition $\sum_{i=1}^{n} \lambda_i \mathbf{p}_i \mathbf{p}_i'$

Extending to general matrices

- In general if $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}' A_n \mathbf{x}$ then $\nabla f(\mathbf{x}) = A_n \mathbf{x}$.
- Assume A_n has spectral decomposition $\sum_{i=1}^{n} \lambda_i \mathbf{p}_i \mathbf{p}_i'$
- Recall that our mean field assumption is $\frac{1}{n}\sum_{i=1}^{n} \lambda_i^2 = o(n)$, which says most eigenvalues are $o(1)$.

Extending to general matrices

- In general if $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}' A_n \mathbf{x}$ then $\nabla f(\mathbf{x}) = A_n \mathbf{x}$.
- Assume A_n has spectral decomposition $\sum_{i=1}^{n} \lambda_i \mathbf{p}_i \mathbf{p}_i'$
- Recall that our mean field assumption is $\frac{1}{n}\sum_{i=1}^{n} \lambda_i^2 = o(n)$, which says most eigenvalues are $o(1)$.
- Thus $A\mathbf{x} \approx \sum_{i=1}^{K} \lambda_i \mathbf{p}_i \mathbf{p}_i' \mathbf{x}$ with $K = o(n)$.

Extending to general matrices

- In general if $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}' A_n \mathbf{x}$ then $\nabla f(\mathbf{x}) = A_n \mathbf{x}$.
- Assume A_n has spectral decomposition $\sum_{i=1}^{n} \lambda_i \mathbf{p}_i \mathbf{p}_i'$
- Recall that our mean field assumption is $\frac{1}{n}\sum_{i=1}^{n} \lambda_i^2 = o(n)$, which says most eigenvalues are $o(1)$.
- Thus $A\mathbf{x} \approx \sum_{i=1}^{K} \lambda_i \mathbf{p}_i \mathbf{p}_i' \mathbf{x}$ with $K = o(n)$.
- This is expressable using at most K bits of information $\{\mathbf{p}_i' \mathbf{x}\}_{i=1}^{K}$, which is why the resulting model is mean field.

Open questions?

- Extend the argument to other exponential families, such as ERGMs.

Open questions?

- Extend the argument to other exponential families, such as ERGMs.
  - Think of edges in graphs as binary random variables.

Open questions?

- Extend the argument to other exponential families, such as ERGMs.
  - Think of edges in graphs as binary random variables.
  - Find for what values of parameters the mean field prediction is tight.

Open questions?

- Extend the argument to other exponential families, such as ERGMs.
  - Think of edges in graphs as binary random variables.
  - Find for what values of parameters the mean field prediction is tight.
  - Partially addressed in Chatterjee-Dembo, results are almost sharp.

Open questions?

- Extend the argument to other exponential families, such as ERGMs.
  - Think of edges in graphs as binary random variables.
  - Find for what values of parameters the mean field prediction is tight.
  - Partially addressed in Chatterjee-Dembo, results are almost sharp.

Open questions?

- Extend the argument to other exponential families, such as ERGMs.
  - Think of edges in graphs as binary random variables.
  - Find for what values of parameters the mean field prediction is tight.
  - Partially addressed in Chatterjee-Dembo, results are almost sharp.
- Solve optimization problem for more graph ensembles.

Thank

You