# Algorithmic Statistics Problem Set 2

Due: 10/15/2025

**Problem 1** Consider a binary Ising model on $n$ nodes with no external field and underlying graph $G$ with edge set $E$. Let its density be given by $p(x) \propto \exp(\sum_{(i,j)\in E} \theta_{ij} x_i x_j)$. Assume that the maximum degree in the graph is $d$ and that every vertex is connected to at least 1 edge. Also assume that $0.1 \leq |\theta_{ij}| \leq 10$ for all $(i,j) \in E$. Prove that there exists a constant $c$ depending only on $d$ (but not $n$) such that every node $x_i$ has a neighbor $x_j$ with $|\mathbb{E}[x_i x_j]| \geq c$.

**Problem 2** Consider a binary Ising model on $n$ nodes with density $p(x) \propto \exp(\sum_{1\leq i<j\leq n} \theta_{ij} x_i x_j)$ (for simplicity, let's consider the case with no external field).

1. Consider the log partition function viewed as a function of $\theta$. Prove that its Hessian $H$ has entries

$$\frac{\partial^2 \log(Z(\theta))}{\partial \theta_{i_1 j_1} \partial \theta_{i_2 j_2}} = \mathbb{E}[x_{i_1} x_{j_1} x_{i_2} x_{j_2}] - \mathbb{E}[x_{i_1} x_{j_1}]\,\mathbb{E}[x_{i_2} x_{j_2}]$$

2. Prove that the Hessian is positive semidefinite (and thus the log partition function is convex)

3. Assume that for all $i$, $\sum_j |\theta_{ij}| \leq \lambda$. Prove that the condition number (maximum singular value divided by minimum singular value) of the Hessian is at most $2^{O(\lambda)} n^{O(1)}$.

4. Imagine getting samples from an unknown Ising model

$$p'(x) \propto \exp\Big(\sum_{1\leq i<j\leq n} \theta'_{ij} x_i x_j\Big)\,.$$

   Deduce that in the infinite sample limit, the negative log likelihood of the samples is a convex function and its unique minimizer is exactly the true parameters i.e. $\theta_{ij} = \theta'_{ij}$ for all $i,j$.

5. This suggests that we might be able to run gradient descent on the objective in (d) to learn an Ising model. It can be shown that with a polynomial number of samples, the objective is still well-behaved and the optimum is close to the true parameters – but there's a catch, and this approach does not lead to a polynomial time algorithm! Why is that? (**Hint:** recall that in general, given $\theta$, it is NP-hard to compute the partition function $Z(\theta)$ and sample from the Ising model with parameters $\theta$)