Recent results in planted assignment problems
Homework 2

1. (Moment Generating Bound) Consider the correlated graph matching model (π, A, B). Let O denote an edge orbit of the edge permutation induced by $\pi^{-1} \circ \widetilde{\pi}$. Let

$$X_O \triangleq \prod_{(i,j) \in O} \frac{\mathcal{P}(B_{\pi(i)\pi(j)} | A_{ij}) \mathcal{P}(B_{\widetilde{\pi}(i)\widetilde{\pi}(j)} | A_{ij})}{\mathcal{Q}(B_{\pi(i)\pi(j)}) \mathcal{Q}(B_{\widetilde{\pi}(i)\widetilde{\pi}(j)})}$$

Prove that

$$\boxed{\mathbb{E}_{\mathcal{Q}}\left[X_O\right] = \begin{cases} \frac{1}{1-\rho^{2|O|}} & \text{Gaussian} \\ 1+\rho^{2|O|} & \text{Erdős-Rényi} \end{cases}}$$

(Hint: check Proposition 1 in https://arxiv.org/abs/2008.10097).

2. (Planted geometric matching model) Consider the following planted geometric matching model proposed in https://arxiv.org/abs/2107.05567. where $Y_i = X_{\pi(i)} + \sigma Z_i$, and X_i, Z_i's are i.i.d. $\mathcal{N}(0, I_d)$ for constant d. The goal is to recover π from the observation of X and Y.

   (a) Prove that when $\sigma = o(n^{-2/d})$, the maximum likelihood estimator succeeds in exact recovery, i.e., $\widehat{\pi}_{\mathrm{ML}} = \pi$ with high probability (Hint: bound the expected total length of "augmenting" cycles using the cycle decomposition of permutation and MGF bound);

   (b) Prove that $\sigma = o(n^{-2/d})$ is necessary for exact recovery.

   (c) Prove that when $\sigma = o(n^{-1/d})$, the maximum likelihood estimator succeeds in almost exact recovery, i.e., $\mathrm{overlap}(\widehat{\pi}_{\mathrm{ML}}, \pi) \to 1$ with high probability;

   (d) Using the mutual information based argument to prove that $\sigma = O(n^{-1+\epsilon}/d)$ is necessary for almost exact recovery;

   (e) Using the posterior sampler based argument to prove that when $\sigma = o(n^{-1/d})$ is necessary for almost exact recovery.

3. (Hypothesis testing in the Planted geometric matching model - open-ended) Consider the hypothesis testing problem proposed in https://arxiv.org/abs/2206.12011, where under H_0, X_i, Y_i are i.i.d. $\mathcal{N}(0, I_d)$ for constant d; under H_1, $Y_i = X_{\pi(i)} + \sigma Z_i$, and X_i, Z_i's are i.i.d. $\mathcal{N}(0, I_d)$. The goal is to detect the hypothesis based on X and Y. Can you determine the sharp detection threshold?