Average-case complexity and statistical inference

Exercise Sheet 1

Please choose some questions below amounting to at least (3) points. Deadline 13th March, email to me `fiona.skerman@math.uu.se` or put a physical copy in my pigeon-hole.

1. (1) Let $SBM(n,p,q)$ be the model constructed as follows. We assume n is even. Let $S^* \in \binom{[n]}{n/2}$, i.e. let S^* be a set of $n/2$ vertices chosen uniformly from all sets of that size in $[n]$. Let $\sigma_v = 1$ if $v \in S^*$ and $\sigma_v = -1$ if $v \notin S^*$. Construct G by choosing each edge to be present independently with probability

$$\mathbb{P}(uv \in E \mid \sigma_u, \sigma_v) = \begin{cases} p & \text{if } \sigma_u \sigma_v = 1 \\ q & \text{otherwise.} \end{cases}$$

   Graph theory notation. For graph $g = ([n], e)$ write $e(g)$ for the number of edges in g. For vertex subset $S \subset [n]$, write $\bar{S} = [n] \backslash S$, write $e_g(S)$ for the number of edges in g with both end points in set S and write $e_g(S, \bar{S})$ for the number of edges in g with one endpoint in S and the other endpoint in $\bar{S}$.

   Let H_1 be $SBM(n,p,q)$. Then show[1]

$$\mathbb{P}_1(G = g) = \binom{n}{n/2}^{-1} \sum_{|S|=n/2} \left(\frac{p(1-q)}{q(1-p)} \right)^{e(g) - e_g(S,\bar{S})} ((1-p)(1-q))^{n^2/4}.$$

2. (1) In the lectures, Lemma 1.1 in the lecture notes, we proved that detection in the matrix model $H_0 = BC'(n,0,0)$ vs $H_1 = BC'(n, k = n^\beta, \lambda = n^{-\alpha})$ is EASY for fixed $\alpha, \beta \in (0,1)$ provided $\beta > \alpha/2 + 1/2$.
   Extend this proof from the BC' model where *exactly* k indices are chosen to be in the community to the BC model where each index is chosen to be in the community with probability k/n independently (and hence we likely get around k indices in the community).

3. (1) Recall the likehood ratio between H_0 and H_1 is defined to be $L(x) = \mathbb{P}_0(X = x)/\mathbb{P}_1(X = x)$. Prove that $\mathbb{E}_0[L(X)] = 1$ and $\mathbb{E}_0[|L_0(X) - 1|] = \sum_x |\mathbb{P}_1(X = x) - \mathbb{P}_0(X = x)|$.

4. (2) **Simulations** Consider the planted dense subgraph model $PDS'(n,p,q)$ as defined in the lecture notes. It is important that we consider the uniform model, i.e. where the set of community vertices S^* has size exactly k and is chosen uniformly at random from all $\binom{n}{k}$ subsets of size k. We want to simulate this graph, then use 3 different algorithms to attempt recovery and plot rate of success of each for different parameter values. A good measure of success is the overlap, if $\hat{S}$ is the k-vertex set returned by the algorithm then the overlap $o(S^*, \hat{S}) = |S^* \cap \hat{S}|/k$.

   A possible set of three algorithms to compare are $\hat{S}_1$ the k vertices of highest degree, $\hat{S}_2$ the iteration of this, see Jiaming Xu lecture notes, and $\hat{S}_3$ a re-iteration of $\hat{S}_2$. A possible set of parameters would be ($n = 100$, $k = 10$, $q = 0.4$, $p = 0.5, 0.6, 0.7, 0.8$) but you may have to play with these a little to get some interesting behaviour.

5. (1) Find a gap in your knowledge and write about it. It could be a detail skipped in the lecture or reading a section of the lecture notes of Lugosi or Wu and Xu and explaining it in your own words (and equations), or something else.

[1] or similar, typos expected.