36-709, Spring 2019
Homework 3

Due Friday, March 8 by 5:00pm in JaeHyeok's mailbox

1. A random matrix A of dimension $n \times m$ is sub-Gaussian with parameter σ^2, written as $A \in SG_{m,n}(\sigma^2)$, when $y^\top Ax$ is $SG(\sigma^2)$ for any $y \in \mathbb{S}^{n-1}$ and $x \in \mathbb{S}^{m-1}$. You may assume that $\mathbb{E}[A] = 0$ (or otherwise replace A by $A - \mathbb{E}[A]$).
   (a) Suppose that the entries of A are independent variables that are $SG(\sigma^2)$. Show that $A \in SG_{m,n}(\sigma^2)$.
   (b) Let $A \in SG_{n,m}(\sigma^2)$ and recall that the operator norm of A is
   $$\|A\|_{\text{op}} = \max_{x \in \mathbb{R}^m, x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{y \in \mathbb{S}^{n-1}, x \in \mathbb{S}^{m-1}} y^\top Ax.$$
   Show that, for some $C > 0$,
   $$\mathbb{E}\left[\|A\|_{\text{op}}\right] \leq C\left(\sqrt{n} + \sqrt{m}\right).$$
   (c) Find a concentration inequality for $\|A\|_{\text{op}}$.

   Hint: work with a 1/4 net for $\mathbb{S}^{n-1}$ and a 1/4 net for $\mathbb{S}^{m-1}$.
2. Exercise 6.10.
3. 6.15 a)
4. Exercise 5.1.
5. Exercise 5.2.
6. Let A be a $n \times n$ symmetric matrix with zero diagonal and off-diagonal entries consisting of $\binom{n}{2}$ independent Bernoulli's. Specifically, for any $i < j$,
   $$A_{i,j} = A_{j,i} \sim \text{Bernoulli}(p_{i,j}),$$
   where each $p_{i,j} \in [0, 1]$. Then A is the adjacency matrix of an *inhomogeneous Bernoulli network*, a random simple graph whose edges are independent Bernoulli's. In particular, if $p_{i,j} = p$ for all $i < j$, A is the adjacency matrix of an Erdö-Renyi random graph.

   In many problems – for example when analyzing the performance of spectral clustering algorithms for community detection – we need a high probability bound for the quantity
   $$\|A - \mathbb{E}[A]\|_{\text{op}}$$
   Let $\alpha = \max_{i<j} p_{i,j}$ and assume that $\alpha = \alpha_n$ is allowed to vanish with n in such a way that $\alpha_n = C_1 \frac{\log n}{n}$, for some C_1. Notice that $\alpha_n n$ is a bound on the maximal degree of the graph.

   Show that there exists a constant C' such that, with probability at least $\frac{1}{n}$,
   $$\|A - \mathbb{E}[A]\|_{\text{op}} \leq \sqrt{C' n \alpha_n \log n}.$$
   Thus, as long as α_n is of larger order than $\frac{\log n}{n}$ (so that the graph may be *sparse,* in the sense that the maximal degree is of smaller order than n), $\|A - \mathbb{E}[P]\|_{\text{op}}$ converges in probability to zero.

Write $A - \mathbb{E}[P] = \sum_{i<j}(A_{i,j} - p_{i,j})(E^{(i,j)} + E^{(j,i)})$*, where* $E^{(i,j)}$ *is the* $n \times n$ *matrix whose entries are all zeros, except for the* (i,j)*th entry, which is* 1*. Use Bernsteion matrix inequality.*

For the current state-of-the art on bounds for this type of problems see, *Afonso S. Bandeira and Ramon van Handel, (2016). Sharp nonasymptotic bounds on the norm of random matrices with independent entries, Ann. Probab. Volume 44, Number 4, 2479-2506.*

7. Let $X_1, \ldots, X_n$ independent and identically distributed sequence taking values in a finite set of cardinality m which, without loss of generality, we may take to be $[m] := \{1, \ldots, m\}$. For each $j \in [m]$, let $\widehat{p}(j) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}(X_i = j)$ be the fraction of the sample points equal to j. Then $(\widehat{p}(1), \ldots, \widehat{p}(m))$ is a probability distribution. Similarly, for each $j \in [m]$, let $p(j) = \mathbb{P}(X_1 = j)$.

   We are interested in bounding the L_1 or total variation distance[1] between the empirical probabilities and the true one, i.e. in bounding

$$\sum_{j=1}^m |\widehat{p}(j) - p(j)|.$$

   (a) Use Hoeffding's inequality and the union bound to show that, with probability at least $1 - \delta$,

$$\sum_{j=1}^m |\widehat{p}(j) - p(j)| \leq m\sqrt{\frac{\log(2m/\delta)}{2n}}.$$

   (b) Here is a way to derive a stronger inequality. Below we will use these standard facts from duality, which give a variational representation of the L_1 and L_∞ vector norms:

$$\|x\|_1 = \sup_{y\colon \|y\|_\infty \leq 1} x^\top y \quad \text{and} \quad \|x\|_\infty = \sup_{y\colon \|y\|_1 \leq 1} x^\top y.$$

   (Recall that $\|x\|_\infty = \max_i |x_i|$ and $\|x\|_1 = \sum_i |x_i|$). Let $\mathcal{N}_\epsilon$ be a ϵ-net of $[-1,1]^m$ with respect to the metric induced by the $\|\cdot\|_\infty$ norm (of course, $0 < \epsilon < 1$). Treat the probability distributions $\widehat{p} := (\widehat{p}(1), \ldots, \widehat{p}(m))$ and $p := (p(1), \ldots, p(m))$ as vectors in $\mathbb{R}^m$.
   Let $x_* \in [-1,1]^m$ be such that $\sum_{j=1}^m |\widehat{p}(j) - p(j)| = (\widehat{p} - p)^\top x_*$. Let $s_* = s_*(x_*)$ be the closest point in $\mathcal{N}_\epsilon$ to x_* (break ties arbitrarily). Of course, both x_* and s_* are random.

      i. Show that

$$\sum_{j=1}^m |\widehat{p}(j) - p(j)| \leq \frac{1}{1-\epsilon}(\widehat{p} - p)^\top s_*$$

      Hint: use the standard fact, from duality, that $\|x\|_1 = \sup_{y\colon \|y\|_\infty \leq 1} x^\top y$ *and Hölder's inequality, which gives that* $|x^\top y| \leq \|x\|_1 \|y\|_\infty$

      ii. Use Hoeffding's inequality and the union bound to bound the the right hand side of the previous expression with probability at least $1 - \delta$. *Hint: it is crucial that you realize that you cannot use Hoeffding inequality on* $(\widehat{p} - p)^\top s_*$ *because each coordinate of* s_* *depends on all the* X_i*'s. Instead you need to use the bound* $(\widehat{p} - p)^\top s_* \leq \max_{y \in \mathcal{N}_\epsilon}(\widehat{p} - p)^\top y$*. Each term* $(\widehat{p} - p)^\top y$ *can be handled with Hoeffding's inequality.*

      iii. At this point, argue that $|\mathcal{N}_\epsilon| \leq \left(\frac{2}{\epsilon}\right)^m$ and pick a value for ϵ. Compare to the naive bound in part (a). You should get a better dependence on m.

   See also the paper: *T. Weissman, E. Ordentlich, G. Seroussi, and S. Verdú. Inequalities for the* ℓ_1 *deviation of the empirical distribution. Technical report, Hewlett-Packard Labs, 2003*

[1] Technically, this is twice the total variation distance.