Math 608D, Assignment 1: Due Friday, Feb 2

1. **Concentration on the cube.** Let $B_\infty^n := \{x \in \boldsymbol{R}^n : \|x\|_\infty \leq 1\} := \{x \in \boldsymbol{R}^n : -1 \leq x_i \leq 1\}$. Let $y = [1, 1, \ldots, 1]/\sqrt{n}$ and consider a diagonal hyperplane through the cube, $H := \{x \in B : \langle x, y\rangle = 0\}$. Let H_ϵ be the ϵ neighborhood of H, i.e., $H_\epsilon := \{x \in B_\infty : |\langle x, y\rangle| \leq \epsilon\}$. Give a bound on

$$\frac{vol(H_\epsilon)}{vol(B_\infty^n)}.$$

   Above, *vol* is the usual Lesbegue measure of volume, i.e., $vol(B_\infty) = 2^n$.

   Note that for $\epsilon \geq \sqrt{n}$, $H_\epsilon = B_\infty^n$. Is the bound you give showing concentration? I.e., does it show that H_ϵ contains most of the volume even when $\epsilon << \sqrt{n}$?

2. Prove the following version of Bernstein inequality:

   Theorem 0.1 (Bernstein inequality for bounded random variables) *Let $X_1, X_2, \ldots, X_N$ be independent, mean-zero, random variables which are all uniformly bounded by a positive scalar M, i.e., $\|X_i\|_\infty \leq M$. Then for any $t > 0$,*

$$P\left(\left|\sum_{i=1}^N X_i\right| \geq t\right) \leq 2\exp\left(-C\min\left(\frac{t^2}{\sum_i \boldsymbol{E}[X_i^2]}, \frac{t}{M}\right)\right).$$

   Remark 0.2 *This version is quite useful for bounded random variables which have standard deviation much lower than M. As seen in class for sums of sub-exponential random variables, there is a combination of sub-Gaussian behavior and sub-exponential behavior in the bound.*

3. As can be seen from the above version of Bernstein inequality, being able to control a random variable in multiple ways can lead to better tail bounds. Now suppose you have a sequence of mean-zero independent random variables $X_1, X_2, \ldots X_N$ having the following properties:

   (a) $\|X_i\|_1 := \boldsymbol{E}|X_i| = \eta$.

   (b) $\|X_i\|_2 := \sqrt{\boldsymbol{E}X_i^2} = \sigma$,

   (c) $\|X_i\|_\infty = M$,

   (d) $\|X_i\|_{\Psi_1} = a_1$,

   (e) $\|X_i\|_{\Psi_2} = a_2$.

   (a) What can you say about the ordering of $\eta, \sigma, M, a_1, a_2$? (Which is largest? etc.)

   (b) In the same spirit as Bernstein inequality above, can you give a tail bound using some subset of the above properties? To be interesting, this bound should be a significant improvement over bounds that can be made using a single property, at least for some values of t. *Note: This is a very open-ended problem. However, I know of at least one useful solution which Xiaowei Li and Halyun Jeong came up with last year, which is leading to a nice improvement in a random matrix theory result.*

4. We show that Gaussian concentration extends, in an important special case, to sub-Gaussian variables. Let X be a vector with independent random entries $X_1, X_2, \ldots, X_n$. Suppose for each i

$$\boldsymbol{E}X_i = 0, \qquad \boldsymbol{E}X_i^2 = 1, \qquad \|X_i\|_{\Psi_2} \leq 10.$$

   Show that the following hold:

(a) $\|\|X\|_2\|_{\Psi_2} \leq C\sqrt{n}$. (This part is not so exciting.)

(b) $\|\|X\|_2 - \sqrt{n}\|_{\Psi_2} \leq C$. (This part matches Gaussian concentration, is tricky to prove, and was the first step in a publication that arose from this class two years ago.)

5. A few months ago Nick Harvey nicely synthesized the analysis of stochastic gradient descent to a toy problem about a certain stochastic process. Prove or disprove Conjecture 2.1 from his notes. To be clear about the conjecture, the inequality should hold for all $\delta < 1/2$.