Problem Set 1

- This problem set is due on **October 2, 2020** in the class.
- Each problem carries 10 points.
- No collaboration allowed. Any two or more identical or nearly-identical solutions will automatically receive zero points each.

1. **(Boosting randomized algorithms)** Imagine we have an algorithm for solving some decision problem (*e.g.,* is a given number p a prime?). Suppose that the algorithm makes a decision at random and returns the correct answer with probability $\frac{1}{2}+\delta$, for some $\delta > 0$, which is just a bit better than a random guess. To improve the performance, we run the algorithm N times and take the majority vote. Show that, for any $\epsilon \in (0,1)$, the answer is correct with probability $1-\epsilon$, as long as $N \geq (1/2)\delta^{-2}\ln(\epsilon^{-1})$.

2. **(Concentration of the L_2-norm of a random Gaussian vector)** Let $X = (X_1, X_2, \ldots, X_n)$ be a random vector with independent standard Gaussian components X_i. For some universal constants k_1, k_2 independent of n, show that

   (a) $\sqrt{n} - k_1 \leq \mathbb{E}||X||_2 \leq \sqrt{n} + k_1$.

   (b) $\text{Var}(||X||_2) \leq k_2$.

3. **(Concentration of Spin Glasses)** Consider a set of $d \geq 2$ binary random variables $\{X_1, X_2, \ldots, X_d\}$, each taking value either $+1$ (*up-spin*) or -1 (*down-spin*), arranged in the form of a complete graph. The weight of the edge (i,j) is denoted by the real number $\theta_{ij}, \forall i < j$. The joint probability distribution of the random variables is given by the following probability mass function

$$\mathbb{P}_\theta(x_1, x_2, \ldots, x_d) = \exp\left(\frac{1}{\sqrt{d}}\big(\sum_{i<j}\theta_{ij}x_ix_j\big) - F_d(\theta)\right), \tag{1}$$

   where the function $F_d : \mathbb{R}^{\binom{d}{2}} \to \mathbb{R}$, known as the *free energy*, is given by

$$F_d(\boldsymbol{\theta}) = \log\left(\sum_{\boldsymbol{x}\in\{\pm1\}^d}\exp\big(\frac{1}{\sqrt{d}}\sum_{i<j}\theta_{ij}x_ix_j\big)\right).$$

   serves to normalize the distribution. The probability distribution (1) was originally used to describe the behaviour of magnets in statistical physics, in which context it is known as the *Ising model*. This model is now extensively used in statistical inference. In this problem, we will study concentration of the free energy function $F_d(\boldsymbol{\theta})$.

(a) Show that $F_d(\cdot)$ is a convex function.

(b) For any two vectors $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \mathbb{R}^{\binom{d}{2}}$, show that $|F_d(\boldsymbol{\theta}) - F_d(\boldsymbol{\theta}')| \leq \sqrt{d/2}||\boldsymbol{\theta} - \boldsymbol{\theta}'||_2$.

(c) Now suppose that the weights $\boldsymbol{\theta}$ are chosen as i.i.d. random variables, so that Eqn. (1) now describes a random family of probability distributions, known as the *Sherrington-Kirkpatrick* (SK) model in statistical physics. As a special case, suppose that the weights are chosen in an i.i.d. manner as $\theta_{ij} \sim \mathcal{N}(0, \sigma^2)$ for each $i \neq j$. Show that

$$\mathbb{P}\left[\frac{F_d(\boldsymbol{\theta})}{d} \geq \log 2 + \frac{\sigma^2}{4} + t\right] \leq \exp(-dt^2/\sigma^2), \quad \forall t > 0.$$

4. **(Balls-in-Bins)** Suppose that T balls are thrown independently and uniformly at random into $2C$ bins. Sort the bins according to the number of balls inside. Let the random variable $M_C(T)$ denote the number of balls in the most occupied C bins among the $2C$ bins.

(a) Using Massart's lemma studied in the class, show that

$$\mathbb{E}(M_C(T)) \leq \frac{T}{2} + \sqrt{CT}.$$

(b) Show that

$$\mathbb{E}(M_C(T)) \geq \frac{T}{2} + \sqrt{\frac{CT}{2\pi}} - O(\frac{1}{\sqrt{T}}).$$

In the above, $O(\cdot)$ denotes the usual asymptotic "Big O" notation.
Hint: Work out the $C = 1$ case first and then extend the result to any arbitrary C.

5. **(Predicting Binary Sequences)** Consider the binary sequence prediction problem studied in the class. Suppose you have side information that the sequence $y_1, y_2, \ldots, y_n$ you will encounter can be partitioned into k parts with an imbalance of $0'$s and 1's within each part, but the endpoints of the segments are not known a priori. How can you leverage this information to get a better prediction method? Design a function ϕ_n that captures this prior knowledge for the best possible k-partition and prove that there exists an algorithm with overall prediction accuracy bounded by this function.