High Dimensional Statistics	**2nd Semester, 2021-2022**

Homework 3 (Deadline: Jun 19)

1. (15 pts) Let $A \in \mathbb{R}^{m\times n}$ be a random matrix whose entries are i.i.d, mean zero, sub-Gaussian entries with parameter σ^2. Show that

$$\|A\|_2 \lesssim C\sigma(\sqrt{m}+\sqrt{n}+t)$$

with probability $1-2e^{-t^2}$, where $C>0$ is some proper constant.

2. (10 pts) Prove the following results.

   - Assume $e^x \le 1+x+x^2$ for $|x|\le b$. Then there holds

   $$e^X \preceq I + X + X^2 \text{ for } \|X\|_2 \le b,$$

   where $X\in\mathbb{R}^{n\times n}$ is a symmetric matrix.
   - Show that $e^{\lambda_{\max}(X)} = \lambda_{\max}(e^X)$, where $X\in\mathbb{R}^{n\times n}$ is a symmetric matrix.

3. (10 pts) In Lecture 8, we have proved the Matrix Bernstein inequality (Theorem 8.16) based on the Lieb inequality (Lemma 8.14). In this problem, you are requested to repeat the proof but using the Golden-Thompson inequality (Lemma 8.13) for tensorization instead of the the Lieb inequality. (**Note**: the concentration result may not be exactly the same.)

4. (5 pts) Show that the RIP property in Definition 8.23 (Eq. (8.9)) is equivalent to

$$\max_{A_S}\|A_SA_S^T - I_s\|_2 \le \delta_s,$$

where A_S denotes any sub-matrix formed by at most s columns of A.

5. (5 pts) Recall the definitions of the TV and Hellinger distances in Lecture 9. Show that for two probability measures $\mathbb{P}$ and $\mathbb{Q}$,

$$\|\mathbb{P}-\mathbb{Q}\|_{\mathrm{TV}} \le H(\mathbb{P}\|\mathbb{Q})\sqrt{1-\frac{H^2(\mathbb{P}\|\mathbb{Q})}{4}}.$$

6. (10 pts) Show the convex property of KL divergence, i.e., prove that for $0\le\alpha\le 1$, we have

   (a) $D(\alpha\mathbb{P}_1+(1-\alpha)\mathbb{P}_2\|\mathbb{Q}) \le \alpha D(\mathbb{P}_1\|\mathbb{Q}) + (1-\alpha)D(\mathbb{P}_2\|\mathbb{Q})$,

   (b) $D(\mathbb{P}\|\alpha\mathbb{Q}_1+(1-\alpha)\mathbb{Q}_2) \le \alpha D(\mathbb{P}\|\mathbb{Q}_1) + (1-\alpha)D(\mathbb{P}\|\mathbb{Q}_2)$.

7. (15 pts) Assume X obeys the uniform distribution on $[\theta,\theta+1]$ and the task is to estimate θ from i.i.d observations $X_1,\cdots,X_n$. A natural estimator is the first order statistic

$$X^{(1)} = \min_k X_k.$$

(a) Prove that

$$\mathbb{E}\left[(X^{(1)} - \theta)^2\right] = \frac{2}{(n+1)(n+2)}.$$

(b) Use Le Cam method to show that the minimax risk to estimate θ in the squared error is lower bounded by c/n^2 where $c > 0$ is a numerical constant.

8. (25 pts) Show the following properties about entropy, conditional entropy and mutual information:

   (a) $H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$;

   (b) $I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$;

   (c) $H(X|Y) \leq H(X)$, $H(Y|X) \leq H(Y)$;

   (d) $H(Y|X) = 0$ if $Y = f(X)$, i.e., when Y is a function of X;

   (e) If X and Y are independent, then $H(X|Y) = H(X)$ and $H(Y|X) = H(Y)$.