High Dimensional Probability and Statistics **2nd Semester, 2023-2024**

Homework 2 (Deadline: May 5)

1. (10 pts) Consider a random variable X taking values in $\mathbb{R}$ with pdf of the form
$$p_\theta(x) = h(x)e^{\langle \theta, T(x)\rangle - \phi(\theta)},$$
where $\theta \in \mathbb{R}^d$. Assume $\nabla\phi(\theta)$ is L-Lipschitz, i.e.,
$$\|\nabla\phi(\theta_1) - \nabla\phi(\theta_2)\|_2 \le L\|\theta_1 - \theta_2\|_2.$$
For a fixed unit-norm vector $v \in \mathbb{R}^d$, show that the random variable $Z = \langle v, T(X)\rangle$ is sub-Gaussian.

2. (10 pts) Compute the KL divergence between two multivariate Gaussian distributions $\mathcal{N}(\mu_1, \Sigma_1)$ and $\mathcal{N}(\mu_2, \Sigma_2)$, where $\mu_1, \mu \in \mathbb{R}^d$ are the means and $\Sigma_1, \Sigma_2 \in \mathbb{R}^{d\times d}$ are two covariance matrices which are symmetric and positive definite.

3. Recall the definition of W_2 distance between two probability distribution μ and ν as follows:
$$W_2(\mu,\nu) = \inf_{X\sim\mu, Y\sim\nu} \left(\mathbb{E}[\|X-Y\|_2^2]\right)^{1/2}.$$
In this problem, we are going to compute the W_2 distance between two Normal distributions $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$, where $\Sigma_1,\ \Sigma_2 \in \mathbb{R}^{n\times n}$ are symmetric and positive.

   - (5 pts) Show that, for any mean zero (X, Y) with covariance matrix
   $$\begin{bmatrix} \Sigma_1 & C \\ C^T & \Sigma_2 \end{bmatrix},$$
   one has
   $$\mathbb{E}[\|X-Y\|_2^2] = \operatorname{trace}(\Sigma_1) + \operatorname{trace}(\Sigma_2) - 2\langle I, C\rangle,$$
   where I is an identity matrix.
   - (5 pts) Show that
   $$\begin{bmatrix} \Sigma_1 & C \\ C^T & \Sigma_2 \end{bmatrix} \succeq 0 \quad \text{if and only if} \quad \Sigma_1 - C\Sigma_2^{-1}C^T \succeq 0.$$
   - (10 pts) Show that the solution to the optimization problem
   $$\max\langle I, C\rangle \quad \text{subject to} \quad \Sigma_1 - C\Sigma_2^{-1}C^T \succeq 0$$
   is given by $\operatorname{trace}\left((\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2})^{1/2}\right)$.
   - (10 pts) Show that
   $$W_2(N(\mu_1,\Sigma_1), N(\mu_2,\Sigma_2))^2 = \|m_1 - m_2\|_2^2 + \operatorname{trace}(\Sigma_1 + \Sigma_2 - 2(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2})^{1/2}).$$

4. (10 pts) Let $A \in \mathbb{R}^{d\times d}$ be a random matrix with each entry being i.i.d bounded random variables in $[a, b]$. Show that the spectral norm of A (i.e., $\|A\|_2$) is sub-Gaussian.

5. (10 pts) Suppose T is a finite set and $\log \mathbb{E}\left[e^{\lambda X_t}\right] \leq \psi(\lambda)$ for all $\lambda \geq 0$ and $t \in T$, where $\psi(\cdot)$ is a convex function and $\psi(0) = \psi'(0) = 0$. Show that

$$\mathbb{E}\left[\max_{t\in T} X_t\right] \leq \psi^{*-1}(\log |T|),$$

where $\psi^*(x) = \sup_{\lambda\geq 0}\{\lambda x - \psi(\lambda)\}$.

6. (10 pts) Let $W \in \mathbb{R}^{m\times n}$ be a random matrix with i.i.d $\mathcal{N}(0, 1)$ entries. We have established an upper bound of $\mathbb{E}\left[\|W\|_2\right]$ using the one-step discretization bound. In this problem you are required to establish an upper bound of $\mathbb{E}\left[\|W\|_2\right]$ using the Dudley integral.