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

# Homework 1 (Deadline: April 20)

1. (10 pts) Let $X$ be mean zero random variable. Show that for any $t > 0$,

$$\mathbb{P}[X > t] \leq \inf_{k=0,1,\cdots} \frac{\mathbb{E}\left[X_+^k\right]}{t^k} \leq \inf_{\lambda \geq 0} \frac{\mathbb{E}\left[e^{\lambda X}\right]}{e^{\lambda t}}, \quad \text{where } X_+ = \max\{X, 0\}.$$

This result basically shows that the tail bound by the moment method is at least as good as the one obtained by the moment generating function (i.e., the Chernoff bound). However, the moment generating function is easy to manipulate as it is a continuous function with respect to $\lambda$.

2. (10 pts) Let $X_1, \cdots, X_n$ be i.i.d samples drawn from a pdf $f(x)$ on the real line. A standard way to estimate $f$ from the samples is the kernel density estimator,

$$\widehat{f}_n(x) = \frac{1}{nh} \sum_{k=1}^{n} K\left(\frac{x - X_k}{h}\right),$$

where $K : \mathbb{R} \to [0, \infty)$ is a kernel function satisfying $\int_{-\infty}^{\infty} K(x) = 1$, and $h > 0$ is a bandwidth parameter. Suppose we evaluate the quality of $\widehat{f}_n(x)$ using the $L_1$ norm

$$\|\widehat{f}_n - f\|_1 := \int_{-\infty}^{\infty} |\widehat{f}_n(x) - f(x)| dx.$$

Show that

$$\mathbb{P}\left[\left|\|\widehat{f}_n - f\|_1 - \mathbb{E}\left[\|\widehat{f}_n - f\|_1\right]\right| \geq t\right] \leq 2e^{-\frac{nt^2}{2}}.$$

3. (10 pts) Let $X$ be sub-Gaussian with parameter $\sigma > 0$. Let $f(x)$ be a Lipschitz function with constant $L > 0$, i.e., $|f(x) - f(y)| \leq L|x - y|$ for all $x, y$. Show that there exists a numerical constant $c > 0$ (which does not depend on any parameter, i.e., universal or absolute constant) such that $f(X)$ is sub-Gaussian with parameter $cL\sigma$. **Hint:** Use alternative characterizations for sub-Gaussian random variables in Lecture 1.2.2.

4. (10 pts) In Lecture 2, we have shown that if

$$\text{Ent}\left[e^{\lambda X}\right] \lesssim \lambda^2 \nu^2 \mathbb{E}\left[e^{\lambda X}\right] \quad \text{for all} \quad \lambda \in \mathbb{R}. \tag{1.1}$$

then $X$ is sub-Gassuain with parameter $\nu$. We have also given two examples (Gaussian and bounded random variables) such that (1.1) holds. This question asks you show that (1.1) holds for all the sub-Gaussian random variables. More precisely, show that if $X$ is $\frac{\nu^2}{4}$-sub-Gaussian, then (1.1) holds.

5. (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.

6. (10 pts) Verify that the equality in the tensorization property of entropy (Theorem 2.11 of Lecture 2) holds for

$$g(X_1, \cdots, X_n) = \exp\left(\lambda \sum_{k=1}^{n} X_k\right),$$

where $X_1, \cdots, X_n$ are independent.

7. (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.

8. (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.

9. (10 pts) Let $X \sim \mathbb{P}$ be $\nu^2$-sub-Gaussian. Show that $W_1(\mathbb{Q}, \mathbb{P}) \lesssim \sqrt{\nu^2 D(\mathbb{Q}\|\mathbb{P})}$ for all $\mathbb{Q} \ll \mathbb{P}$.

10. (10 pts) Let $a = (a_1, \cdots, a_n)$ and $b = (b_1, \cdots, b_n)$ be two probability measures defines on the discrete space $\mathcal{X} = \{x_1, \cdots, x_n\}$. Assume the metric (transportation cost per unit) has the following form

$$d(x_i, x_j) = \begin{cases} 0 & \text{if } x_i = x_j, \\ 1 & \text{otherwise.} \end{cases}$$

Compute the Wasserstein distance $W_1(a, b)$.

11. (15 pts) Study the following basic questions with respect to functions:

    (a) Suppose $f(x_1, \cdots, x_n)$ is convex with respect to each $x_i, i = 1, \cdots, n$ when the rest of the coordinates are fixed (i.e., separately convex ). Is $f$ a convex function. Prove or disprove by a counter example.

    (b) Let $f(x_1, \cdots, x_n)$ be a smooth function. Assume $f$ is Lipschitz:

    $$|f(x) - f(y)| \le L\|x - y\|_2.$$

    Show that the gradient of $f$, denoted $\nabla f(x)$, satisfies $\|\nabla f(x)\|_2 \le L$.