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

Homework 1 (Deadline: Apr 7)

1. (5 pts) Assume X is ν^2-sub-Gaussian. Show that

$$\mathrm{Var}\left[X\right] \leq \nu^2.$$

2. (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 $cL\sigma$.

3. (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

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

4. (10 pts) Let $A \in \mathbb{R}^{n \times n}$ be a symmetric and positive definite matrix, and let $g \sim \mathcal{N}(0, I_n)$. Try to develop a Bernstein type upper tail for the deviation of $X = g^T A g$ from its mean.

5. (20 pts) Consider a sequence of independent random variables $(X_1, \cdots, X_n)$ that satisfy $0 \leq X_k \leq b$ for each k. Let $Z = \sum_{k=1}^{n} X_k$. Show that

$$\mathbb{P}\left[Z \geq (1+t)\mathbb{E}\left[Z\right]\right] \leq \left(\frac{e^t}{(1+t)^{1+t}}\right)^{\mathbb{E}[Z]/b} \quad \text{for } t > 0,$$

$$\mathbb{P}\left[Z \leq (1-t)\mathbb{E}\left[Z\right]\right] \leq \left(\frac{e^t}{(1-t)^{1-t}}\right)^{\mathbb{E}[Z]/b} \quad \text{for } t \in (0, 1).$$

6. (10 pts) In Lecture 2, we have shown that if

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

then X is sub-Gassuain with parameter ν. 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.