S&DS 602: Homework 6

Due Wednesday, October 9 at 2PM, via Gradescope

1. Let Z be a random variable with finite second moment, and let $\mathbb{M}Z$ be a median of the distribution of Z (i.e. $\mathbb{P}[Z \geq \mathbb{M}Z] \geq 1/2$ and $\mathbb{P}[Z \leq \mathbb{M}Z] \geq 1/2$). Show that

$$|\mathbb{M}Z - \mathbb{E}Z| \leq \mathbb{E}|Z - \mathbb{M}Z| \leq \sqrt{\operatorname{Var} Z}$$

Consequently, show that there exist constants $C, c > 0$ for which

$$\mathbb{P}[|Z - \mathbb{E}Z| \geq t] \leq Ce^{-ct^2} \text{ for all } t \geq 0$$

if and only if there exist constants $C', c' > 0$ for which

$$\mathbb{P}[|Z - \mathbb{M}Z| \geq t] \leq C'e^{-c't^2} \text{ for all } t \geq 0.$$

2. (van Handel 4.9) Let $X \in \{+1, -1\}^n$ have i.i.d. Rademacher coordinates, and define

$$A = \left\{ x \in \{+1, -1\}^n : \sum_{i=1}^{n} x_i \leq 0 \right\}, \qquad f(X) = \min_{x \in A} \|X - x\|_2.$$

(a) Show that $|f(x) - f(y)| \leq \|x - y\|_2$ for all $x, y \in \{+1, -1\}^n$.

(b) Show that there is a constant $c > 0$ for which $\mathbb{P}[f(X) \geq n^{1/4}] > c$ for all large n. [Hint: Show that $f(X) \geq n^{1/4}$ as long as $\sum_{i=1}^{n} X_i \geq \sqrt{n}$.]

(c) Show that $f(X)$ has median $\mathbb{M}f(X) = 0$, and hence $f(X)$ cannot be σ^2-subgaussian for any $\sigma^2 < \infty$ that is independent of the dimension n.

[This example illustrates that the discrete gradient cannot be replaced by the Euclidean gradient ∇ for concentration on the hypercube from Lecture 5, and that the assumption of convexity of f is needed for the convex concentration inequality of Lecture 6.]

3. (van Handel 4.13) (a) Let X be a random variable with law P, and let $f : \mathbb{R} \to \mathbb{R}$ be such that $\mathbb{E}_P[f(X)e^{f(X)}] < \infty$. Define the law Q by $\frac{dQ}{dP}(x) = \frac{e^{f(x)}}{\mathbb{E}_P[e^{f(X)}]}$. Show that

$$\frac{\operatorname{Ent}_P[e^{f(X)}]}{\mathbb{E}_P[e^{f(X)}]} \leq \mathbb{E}_Q f(X) - \mathbb{E}_P f(X).$$

(b) Suppose further that f is convex and continuously differentiable. Show that

$$\frac{\mathrm{Ent}_P[e^{f(X)}]}{\mathbb{E}_P[e^{f(X)}]} \leq \inf_{(X,Y)\sim\text{couplings of }(P,Q)} \mathbb{E}[f'(Y)(Y-X)].$$

If P satisfies the T_2-inequality

$$\inf_{(X,Y)\sim\text{couplings of }(P,Q)} E[(X-Y)^2] \leq 2D_{\mathrm{KL}}(Q\|P),$$

show that this implies the log-Sobolev inequality (for convex f as above)

$$\mathrm{Ent}_P[e^{f(X)}] \leq 2\,\mathbb{E}_P[f'(X)^2 e^{f(X)}].$$

4. Let $X \in \mathbb{R}^{n\times d}$ be a fixed regression design with $\|X\|_{\mathrm{op}} \leq 1$, and let $y = X\theta + \varepsilon$ where

$$\theta = (\theta_1, \ldots, \theta_d), \qquad \varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$$

are independent. Suppose $\theta_1, \ldots, \theta_d$ are i.i.d. variables in $[0,1]$ with density $g(\theta)$, and $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. $\mathcal{N}(0,1)$ errors. Consider the marginal log-likelihood

$$Z = \frac{1}{d}\log\int_{[0,1]^d} \exp\left(-\frac{\|y-Xu\|_2^2}{2}\right)\prod_{j=1}^d g(u_j)du_j$$

as a function of the variables (θ, ε) defining y. Show that for some universal constants $C, c > 0$,

$$\mathbb{P}[|Z - \mathbb{E}Z| \geq t] \leq Ce^{-c\min(t^2,t)} \text{ for all } t \geq 0.$$

[Hint: Write $Z = -\frac{\|y\|_2^2}{2d} + f(y)$ where $f(y)$ is a convex and Lipschitz function of (θ, ε). Bound separately $f(y) - \mathbb{E}[f(y) \mid \varepsilon]$, $\mathbb{E}[f(y) \mid \varepsilon] - \mathbb{E}f(y)$, and $\frac{\|y\|_2^2 - \mathbb{E}\|y\|_2^2}{2d}$.]