Math 281C Homework 6

Due: May 31, 5 pm

1. Let $R_n : \Theta \to \mathbb{R}$ be a sequence of random functions and $R(\theta) = \mathbb{E}R_n(\theta)$. Let $d : \Theta \times \Theta$ be some distance on Θ. Denote $\theta_0 = \operatorname{argmin}_\theta R(\theta)$, and for $0 < \delta < \infty$, define $\Theta_\delta = \{\theta : d(\theta, \theta_0) \le \delta\}$. For $\alpha \in (0, 2), \sigma < \infty$, and $D > 0$, assume we have the continuity bound

$$\mathbb{E}\left[\sup_{\theta \in \Theta_\delta} |(R_n(\theta) - R(\theta)) - (R_n(\theta_0) - R(\theta_0))|\right] \le \frac{\sigma\delta^\alpha}{\sqrt{n}}$$

for all $\delta \le D$. Assume in addition that for some parameters $\beta \in [1, \infty]$ and $v > 0$, we have

$$R(\theta) \ge R(\theta_0) + v d(\theta, \theta_0)^\beta$$

for $d(\theta, \theta_0) \le D$. Let $\widehat{\theta}_n = \operatorname{argmin}_\theta R_n(\theta)$ and assume that $\widehat{\theta}_n$ is consistent for θ_0. Give the largest rate r_n you can for which

$$r_n d(\widehat{\theta}_n - \theta_0) = O_P(1) \quad \text{as} \quad n \to \infty.$$

2. Suppose that we have $X_i \in \mathbb{R}^d$ and observe

$$Y_i = \langle X_i, \theta_0 \rangle + \epsilon_i, \quad \text{where} \quad \epsilon_i = B_i Z_i$$

for $i = 1, \dots, n$. Here $B_i \in \{0, 1\}$ is independent of Z_i and X_i, and $\mathbb{P}(B_i = 0) = p > 1/2$. The variable Z_i has arbitrary distribution, independent of X_i and $\mathbb{E}[|Z_i|] < \infty$. We decide to estimate θ_0 using the absolute loss $\ell(\theta; x, y) = |y - \langle x, \theta \rangle|$. In other words,

$$\widehat{\theta}_n = \operatorname*{argmin}_{\theta \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^n \ell(\theta; X_i, Y_i).$$

Let $R_n(\theta)$ and $R(\theta)$ denote the empirical risk and population risk respectively.

   (a) Show that for any $\theta \in \mathbb{R}^d$, we have

$$R(\theta) - R(\theta_0) \ge (2p - 1)\mathbb{E}[|\langle X, \theta - \theta_0 \rangle|].$$

Now for $v \in \mathbb{S}^{d-1}$, denote $\sigma_v^2 = \mathbb{E}[(\langle v, X \rangle)^2]$, and assume there exist two constants c_1 and c_2 such that

$$\mathbb{P}(|\langle v, X \rangle| \ge c_1 \sigma_v) \ge c_2 > 0.$$

Moreover, assume there is a constant $D < \infty$ such that $\|X\|_2 \le D$ with probability 1, and $\mathbb{E}[XX^\top] = \Sigma > 0$.

   (b) Show that for any $v \in \mathbb{R}^d$,

$$\mathbb{E}[|\langle v, X \rangle|] \ge \rho \|v\|_2,$$

   where ρ is a constant that depends on the distribution of X but is independent of v.

   (c) Show that there exists a constant $\sigma < \infty$ which may depend on (D, d) such that for any $\delta > 0$,

$$\mathbb{E}\left[\sup_{\theta : \|\theta - \theta_0\|_2 \le \delta} |(R_n(\theta) - R(\theta)) - (R_n(\theta_0) - R(\theta_0))|\right] \le \frac{\sigma\delta}{\sqrt{n}}.$$

   (d) At what rate does $\widehat{\theta}_n$ converge to θ_0? You may assume that $\widehat{\theta}_n$ is consistent for θ_0.