STAT260 Problem Set 1

Due February 4th in class

Regular problems:

1. Suppose that p^* has bounded kth moments in the sense that $\mathbb{E}[|X-\mu|^k] \leq \sigma^k$ for some σ. Show that the trimmed mean applied to $\tilde{p}$ has error $\mathcal{O}(\sigma\epsilon^{1-1/k})$ if $\mathsf{TV}(\tilde{p}, p^*) \leq \epsilon$.

2. Call a distribution p over $\mathbb{R}$ (s,ϵ)-*stable* if $\mathbb{P}_{x\sim p}[x \geq \mu + s] < \frac{1}{2} - \epsilon$ and $\mathbb{P}_{x\sim p}[x \leq \mu - s] < \frac{1}{2} - \epsilon$. If p^* is (s,ϵ)-stable, show that the median applied to $\tilde{p}$ estimates the mean with error at most s.

3. Show that a 1-dimensional Gaussian with variance σ^2 is $(\mathcal{O}(\sigma\epsilon), \epsilon)$-stable for $\epsilon \leq \frac{1}{4}$.

4. Call a distribution p over $\mathbb{R}^d$ (s,ϵ)-stable if $\langle X, v\rangle$ is (s,ϵ)-stable for $X \sim p$ and all unit vectors v. Show that the empirical distribution of n samples from p is $(2s, \frac{\epsilon}{2} - \mathcal{O}(\sqrt{d/n}))$-stable around the true mean μ with probability at least $1 - 2\exp(-c\epsilon^2 n)$ for a constant $c > 0$. *[Hint: first show this in 1 dimension, then union bound.]*

5. For a distribution p, define the *Tukey median* $\hat{\theta}_{\mathsf{Tukey}}$ as

$$\hat{\theta}_{\mathsf{Tukey}}(p) = \arg\max_{\theta\in\mathbb{R}^d} D(\theta), \text{ where } D(\theta) = \inf_{v\in\mathbb{R}^d} \mathbb{P}_{X\sim p}[\langle X-\theta, v\rangle \geq 0]. \tag{1}$$

   This is also sometimes called the maximum depth point. Define the *Tukey depth* to be the maximum value of $D(\theta)$.

   Show that if p^* has Tukey depth D and is $(s, c\cdot(\epsilon + (1/2 - D)))$-stable for a sufficiently large constant c, then $\|\hat{\theta}_{\mathsf{Tukey}}(\tilde{p}) - \mu(p)\|_2 \leq s$.

Challenge problems (credit for at most one; turn in as a separate document typset in LaTeX):

6. Let p_n^* be the empirical distribution of n samples from $\mathcal{N}(\mu, \sigma^2 I)$, and suppose that $\mathsf{TV}(\tilde{p}_n, p_n^*) \leq \epsilon$. Show that with high probability, the Tukey median applied to $\tilde{p}_n$ estimates μ with error $\mathcal{O}(\sigma\epsilon)$ as long as ϵ is sufficiently small and $n \gg \frac{d}{\epsilon^2}$. *[You may assume the results of the previous exercises.]*

7. The *geometric median* $\hat{\theta}_{\mathsf{geom}}(p)$ is defined as the minimizer θ of $\mathbb{E}_{X\sim p}[\|X-\theta\|_2]$. Let $p^* = \mathcal{N}(\mu, I)$ and suppose that $\mathsf{TV}(\tilde{p}, p^*) \leq \epsilon$ and the new perturbed points in $\tilde{p}$ are constrained to have norm at most $2\sqrt{d}$. Construct such a $\tilde{p}$ such that $\|\hat{\theta}_{\mathsf{geom}}(\tilde{p}) - \mu(p)\|_2 = \Omega(\epsilon\sqrt{d})$.