Regular problems:

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

2. Call a distribution \( p \) over \( \mathbb{R} \) \((s, \epsilon)\)-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 \((s, \epsilon)\)-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 \( \sigma^2 \) is \((O(\sigma \epsilon), \epsilon)\)-stable for \( \epsilon \leq \frac{1}{4} \).

4. Call a distribution \( p \) over \( \mathbb{R}^d \) \((s, \epsilon)\)-stable if \( \langle X, v \rangle \) is \((s, \epsilon)\)-stable for \( X \sim p \) and all unit vectors \( v \). Show that the empirical distribution of \( n \) samples from \( p \) is \((2s, \epsilon^2 - O(\sqrt{d/n}))\)-stable around the true mean \( \mu \) 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 \( \hat{\theta}_{\text{Tukey}} \) as

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

(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 \( \|\hat{\theta}_{\text{Tukey}}(\tilde{p}) - \mu(p)\|_2 \leq s \).

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 \( TV(\tilde{p}_n, p_n^*) \leq \epsilon \). Show that with high probability, the Tukey median applied to \( \tilde{p}_n \) estimates \( \mu \) with error \( O(\sigma \epsilon) \) as long as \( \epsilon \) is sufficiently small and \( n \gg \frac{d}{\epsilon^2} \). [You may assume the results of the previous exercises.]

7. The geometric median \( \hat{\theta}_{\text{geom}}(p) \) is defined as the minimizer \( \theta \) of \( E_{X \sim p}[\|X - \theta\|_2] \). Let \( p^* = \mathcal{N}(\mu, I) \) and suppose that \( TV(\tilde{p}, p^*) \leq \epsilon \) 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 \( \|\hat{\theta}_{\text{geom}}(\tilde{p}) - \mu(p)\|_2 = \Omega(\epsilon \sqrt{d}) \).