MATH524 – Spring 2025

Problem Set: Week 9

1. Let $X_1, \dots, X_n$ be an i.i.d. sample from $\mathcal{N}(\mu, \sigma^2)$ where σ is a known constant. Prove using Le Cam's two-points lemma that
$$\sup_{\mu \in \mathbb{R}} \mathbb{E}|\tilde{\mu} - \mu| \geq \frac{C}{\sqrt{n}},$$
for any estimator $\tilde{\mu}$ with some constant C.

2. (**Sparse mean-vector estimation**) In this exercise we will use a modified version of the Varshamov-Gilbert construction from the lecture notes to help identify minimax lower-bounds for sparse vector estimation.

   Lemma 1 (Sparse Varshamov-Gilbert)
   Conisder $\Omega = \{\omega \in \{0,1\}^d : \|\omega\|_0 \leq s\}$. Then, there exists $\Omega' \subseteq \Omega$ such that

   - $\|\omega\|_0 = s$ *for every* $\omega \in \Omega'$
   - $H(\omega_i, \omega_j) \geq s/2$
   - $|\Omega'| \geq c(de/s)^s$.

   Now, consider $X \sim \mathcal{N}(\theta, \sigma^2 I_d)$ with $\|\theta\|_0 \leq s$ (i.e. the mean vector is sparse). The following upper bound is well-known:
$$\mathbb{E}[\|\hat{\theta} - \theta\|^2] \precsim \frac{\sigma^2 s \log d}{n}$$
   where
$$\hat{\theta} = \begin{cases} \bar{X}_n & \text{if } \bar{X}_n \gtrsim \sigma\sqrt{\frac{\log d}{n}} \\ 0 & \text{otherwise} \end{cases}$$
   Using Lemma 1 to construct a reasonable packing set and Fano's inequality, show that a matching lower bound (up to log factors) can be established.