STAT 535, Homework 1

Due date: October 3 Thursday 23:59:59. Submit the homework through Canvas in a PDF file.

1. **(10 pts)** Suppose that $X_1, \cdots, X_n \in \mathbb{R}^d$ are IID observations with $\mathbb{E}(X_1) = 0$ and $\|X_1\|_{\max} \leq L$ for some positive number L. Let $\Sigma = \mathbb{E}(X_1 X_1^T)$ be the population covariance and $\widehat{\Sigma}_n = \frac{1}{n}\sum_{i=1}^n X_i X_i^T$ be a modified sample covariance matrix. Show that when $d = d_n \to \infty$ as $n \to \infty$,

$$\|\widehat{\Sigma}_n - \Sigma\|_{\max} = O_P\left(\sqrt{\frac{\log d}{n}}\right),$$

   where $\|\cdot\|_{\max}$ is the matrix max norm.

2. **(10 pts)** Let $X_1, \cdots, X_n$ be IID from a univariate Gaussian $N(0, \sigma^2)$. Show that

$$\max\{X_1, \cdots, X_n\} = O_P\left(\sqrt{\log n}\right).$$

3. **(20 pts)** Suppose that we have a sequence of random variables $X_1, X_2, X_3, \cdots$. Derive the following results.

   (a) **(5 pts)** If there exists positive numbers A_0, A_1, q such that

$$P(|X_n| \geq \epsilon) \leq A_0 e^{-A_1 a_n \epsilon^q}$$

   for all n and a_n is an increasing sequence, then

$$X_n = O_P\left(\frac{1}{{a_n}^{1/q}}\right).$$

   (b) **(5 pts)** If there exists positive numbers q such that

$$P(|X_n| \geq \epsilon) \leq \frac{b_n}{\epsilon^q}$$

   for all n and b_n is an decreasing sequence, then

$$X_n = O_P\left(\frac{1}{{b_n}^{-1/q}}\right).$$

   (c) **(10 pts)** Does the fact that $\mathbb{E}(X_n) = O(a_n)$ implies $X_n = O_P(a_n)$? Does the fact that $X_n = O_P(a_n)$ implies $\mathbb{E}(X_n) = O(a_n)$? If yes, prove it. If no, find a counterexample.

4. **(10 pts)** Let $X_1, \cdots, X_n$ be IID from a multinomial distribution such that $P(X_1 = j) = p_j$ for $j = 1, \cdots, 6$ and $\sum_{j=1}^6 p_j = 1$. The parameters of interest are $p_1, \cdots, p_6$ and the parameter space is $\Theta = \{(p_1, \cdots, p_6) : p_j > 0, \sum_{j=1}^6 p_j = 1\}$.

   (a) **(5 pts)** Find the MLE $\widehat{p}_{\mathsf{MLE}}$.

   (b) **(5 pts)** Suppose we want to test the following hypothesis

$$H_0 : p_1 = p_2 = p_3, \quad p_4 = p_5 = p_6.$$

   Write down the LRT test statistic and the reference distribution under H_0.