1. [30 points] Constructing and interpreting confidence intervals: Suppose that we have $X_1, \ldots, X_n \sim \text{Ber}(p)$, where $p$ is unknown.

   - Use Hoeffding’s bound to construct a 95% confidence interval for $p$.
   - Use the CLT to construct an approximate 95% confidence interval for $p$.
   - Simulate the following setup: Repeat each of the following experiments 10000 times. Vary the sample size from 10 to 10000 on a logarithmic scale, and vary $p$ from 0 to 1 on a linear scale. For each value of $(n, p)$ compute the coverage probability of the CLT interval and the Hoeffding interval, i.e. for each value of $(n, p)$ and for each of the 10000 experiments construct the interval and check if the true parameter is inside the constructed interval.

   Plot and summarize your findings. Particularly, do the two types of intervals have the advertised coverage and are they conservative?

2. [20 points] Finding and Using Pivots:

   - Suppose that $X_1, \ldots, X_n \sim \text{Exp}(\lambda)$, then show that
     $$Q(X_1, \ldots, X_n, \lambda) = \frac{2 \sum_{i=1}^{n} X_i}{\lambda}$$

     is a valid pivot. You can use the fact that the sum of exponentials has a Gamma distribution.

     Use this pivot to construct a $(1 - \alpha)$-confidence interval for $\lambda$.

   - A location family is a collection of shifted (mean 0) distributions, i.e. for some fixed distribution $F$ with mean 0, we define:
     $$\mathcal{P}_\theta = \{ G : G(x) = F(x - \theta) \} .$$

     Show that $\overline{X} - \theta$ is always a pivot for a location family.

3. [15 points] Multiple Testing 1: Suppose that we have $d$ hypotheses, which are ordered a-priori (i.e. without looking at the data). We use the following procedure:
   
   (a) If $p_1 \geq \alpha$, we accept all the null hypotheses and stop, else we reject $H_{01}$ and continue to the next step.

   (b) If $p_2 \geq \alpha$, we accept $H_{02}, \ldots, H_{0d}$ and stop, otherwise we reject $H_{02}$ and continue.

   (c) :
Suppose that the p-values are independent: does this procedure control the FWER at \( \alpha \)? Would this still be the case if the p-values were dependent? Is this procedure more/less/incomparable in power to the Bonferroni procedure? Explain your answer.

4. **[15 points] Multiple Testing 2:** Suppose that a scientist plans to do 1000 gene association tests. Before looking at the data she orders the hypotheses according to her belief, placing “promising” genes first and then less promising ones and so on. She then does the following:

   (a) She computes p-values for every hypothesis.
   
   (b) She rejects the first null hypothesis if the p-value \( \leq \alpha/2 \), the second null if the p-value is \( \leq \alpha/4 \) and so on.

   • Does her procedure control the FWER at a reasonable level? If not, explain why not and if yes, give a proof.
   
   • Does her procedure control the False Discovery Rate (FDR) at a reasonable level? If not, explain why not and if yes, give a proof.

5. **[20 points] Weak \( \ell_q \) sparsity:** In lecture we showed that the hard-thresholding algorithm for the Gaussian Sequence Model achieves the rate,

\[
R(\hat{\theta}, \theta^*) \leq CR\sigma \sqrt{\frac{\log d}{n}},
\]

for estimating a vector \( \theta^* \) that is \( \ell_1 \)-sparse, i.e. satisfies \( \sum_{i=1}^{d} |\theta_i| \leq R \) (where \( C > 0 \) is some constant). Suppose that, \( \theta^* \) is instead \( \ell_q \) sparse for some \( q \in (0, 1] \), i.e.

\[
\sum_{i=1}^{d} |\theta_i|^q \leq R_q.
\]

Then show that for some constant \( C > 0 \) the hard-thresholding estimator has risk:

\[
R(\hat{\theta}, \theta^*) \leq CR_q \left( \frac{\sigma^2 \log d}{n} \right)^{1-\frac{q}{2}}.
\]

Notice that once again the estimator can have risk \( \to 0 \) even when \( d \gg n \). Furthermore, notice that the same hard-thresholding estimator works for a variety of different notions of sparsity.