1. Suppose we wish to test \( n \) normal means \( \mu_i \) from \( X_i \sim \mathcal{N}(\mu_i, 1) \). One way to do this might be as follows: simulate \( n \) iid random variables \( X_i \) from \( \mathcal{N}(0, 1) \) and select \( \tau \) as

\[
\tau = \min \left\{ t : \frac{1 + |\hat{S}(t)|}{1 + |S(t)|} \leq q \right\},
\]

where

\[
\hat{S}(t) = \{ i : |\bar{X}_i| \geq t \text{ and } |\bar{X}_i| > |X_i| \} \\
S(t) = \{ i : |X_i| \geq t \text{ and } |\bar{X}_i| \leq |X_i| \};
\]

then reject those hypotheses in \( S(\tau) \).

(a) Do you expect this to control the FDR at level \( q \)? Explain why or why not.

(b) Take \( n = 1,000 \) and simulate the FDR and power of the above method in the following two settings: (1) 80% of the \( X_i \)'s are \( \mathcal{N}(0, 1) \) and 20% are \( \mathcal{N}(5, 1) \); (2) 80% of the \( X_i \)'s are \( \mathcal{N}(0, 1) \) and 20% are \( \mathcal{N}(2, 1) \). Compare FDR and power with BHq and comment on your findings.

(c) Same as (b) but with 95% of nulls instead.

2. Suppose I wish to test whether a given regression coefficient in a logistic model is null or not.

(a) Explain you would compute p-values classically (for instance, explain how R computes p-values). Are these classical p-values truly uniform under the null? If not, when are they approximately uniform?

(b) Imagine we have \( n \) samples of the form \((X_i, Y_i)\), where \( X_i \sim \mathcal{N}(0, I_p) \) and \( Y_i = \pm 1 \) with probability \( 1/2 \), independently from \( X \). That is, the covariates are i.i.d. \( \mathcal{N}(0, 1) \) and we are under the global null. For \( n = 1500 \) and \( p = 500 \), plot histograms of the empirical distribution of the p-values. Comment on your findings.

(c) Repeat (b) for various values of the ratio \( p/n \). What do you get when \( p/n > 1/2 \)? Can you explain this phenomenon?

(d) What would happen if ou were to use classical p-values for multiple testing when \( p/n = 0.3 \), say?

(e) We are in the setup of (b) but now have a fraction of nonzero regression coefficients. Does the distribution of a null p-value seem to change (compared to what it is under the global null)? Explain your answer.

3. Consider the proof from Lecture 10, Section 3, and recall the ordering \( x \geq y \text{ iff } x_i \geq y_i \text{ for all } i \). Show that if \( f : \mathbb{R}^d \to \mathbb{R} \) is a non-decreasing function meaning that \( x \geq y \implies f(x) \geq f(y) \), then for each \( t_1 \leq t_1' \),

\[
\mathbb{E}(f(X)|X_1 = t_1) \leq \mathbb{E}(f(X)|X_1 = t_1').
\]