1. Suppose we wish to test \( n \) hypotheses \( H_1, H_2, \ldots, H_n \). As usual we assume that the null p-values are uniformly distributed. In this problem, we are interested in procedures which operate in two steps:

**Step 1** Select a set \( S \subset \{1, \ldots, n\} \) of ‘promising’ hypotheses.

**Step 2** Apply a multiple testing procedure to test those hypotheses in \( S \), namely, \( \{H_i\}_{i \in S} \).

Below we shall assume that the selection step is monotone in the following sense: if \( S(p) \) is the set of selected hypotheses on the basis of the \( n \) p-values \((p_1, \ldots, p_n)\), then \( p_i \leq p'_i \) for all \( i \) (\( p \leq p' \) for short) implies that \( S(p') \subset S(p) \).

(a) Suppose we apply the Benjamini-Hochberg (BH) procedure to the selected set of hypotheses with an FDR target level set to \( q \) (this means that the critical thresholds would be equal to \( q_i/|S| \) for \( i = 1, 2, \ldots, |S| \)). Under independence of all \( n \) p-values, would you expect FDR control at level \( q \)? Explain why or why not. Similarly, imagine you were to apply the Bonferroni correction at level \( \alpha/|S| \), would you expect FWER control at level \( \alpha \)?

(b) Suppose now that you apply the BH procedure to the selected hypotheses with an FDR target set to \( q|S|/n \). Under independence between all the p-values, show that this two-step procedure would control the FDR at level \( q \).

*Hint:* You may use the following claim: whenever a function \( f : (p_1, \ldots, p_n) \to [0,1] \) is non-increasing (recall that this means that \( p \leq p' \) implies \( f(p) \geq f(p') \)), we have

\[
\mathbb{E} \left[ \frac{1(p_i < f(p))}{f(p)} \right] \leq 1,
\]

provided the p-values obey the PRDS property.

(c) Suppose then that the \( n \) p-values actually obey the PRDS property, would FDR control at level \( q \) continue to hold? Explain why or why not.

(d) Under independence between the p-values, can I set a nominal threshold higher than \( q|S|/n \) and expect FDR control in general? Explain why or why not.

(e) Describe an application where it might make sense to use the two-step procedure we have just described.

(f) *Bonus question:* Prove the claim from the hint (you will get extra points if you do this).

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

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

where

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

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

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

1
(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.