Statistical Inference II - PhD ECOSTAT - University of Milano-Bicocca

Homework 4

To submit via e-mail by May 15, 2020 h 14:00.

1 Multiple testing

Consider a collection of $m = 100$ null hypotheses. Assume that 90 of them are true, while the the remaining 10 are false. Assume further that p-values from true hypotheses are i.i.d. Uniform$(0, 1)$ and, independently, p-values from false hypotheses are i.i.d. Uniform$(0, 0.2)$. Suppose to reject all the hypotheses for which the corresponding p-value is less or equal than 0.1. What is

1. the expected number of rejected hypotheses;
2. the familywise error rate;
3. the false discovery rate (by using simulations);
4. the false discovery rate of the Benjamini-Hochberg procedure at level $\alpha = 0.1$.

2 Holm

Prove that Holm's procedure controls the familywise error rate at α.

3 Two-step procedure

Suppose we wish to test $m > 1$ hypotheses $\mathcal{H} = \{H_1, \ldots, H_m\}$ by using Bonferroni and Benjamini-Hochberg procedures that operate in two-step:

Step 1: Specify $\alpha \in (0, 1)$. Select the "interesting" hypotheses

$$\mathcal{S} = \{H_i \in \mathcal{H} : p_i \leq \alpha\}$$

Step 2: Apply Bonferroni and Benjamini-Hochberg procedures at level α to the selected hypotheses in $\mathcal{S}$ as if they were given a priori (e.g. Bonferroni rejection set is $\mathcal{R} = \{H_i \in \mathcal{S} : p_i \leq \frac{\alpha}{|\mathcal{S}|}\}$)

Calculate the FWER of this Bonferroni two-step procedure and the FDR of this Benjamini-Hochberg two-step procedure by assuming that all m hypotheses are true and $p_1, \ldots, p_m$ are i.i.d. Uniform(0,1).

4 Conservative BH

Consider a multiple testing procedure that rejects less than Benjamini-Hochberg: let $p_{(i)}$ be the ordered p-values with corresponding hypothesis $H_{(i)}$, then this procedure rejects $H_{(1)}, \ldots, H_{(i_0)}$ where $i_0 \leq i_{\mathrm{BH}}$ with

$$i_{\mathrm{BH}} = \max\{i \in \{1, \ldots, m\} : p_{(i)} \leq \frac{i\alpha}{m}\}$$

Assuming independence of p-values, does this procedure control FDR at α? Explain why or why not.