36-743 Fall 2019

(Some) Statistical Aspects of Reproducibility

The one and only homework

Lecturer : Aaditya Ramdas

Question 1 (Power of Bonferroni under dependence) *Bonferroni controls the Type-1 error (global null) under arbitrary dependence between the p-values, and also controls the FWER/PFER under arbitrary dependence. This question is about the achieved versus targeted FWER. Does Bonferroni come closer to using up its error budget for independent p-values, or for very dependent p-values? Provide a well-reasoned mathematical argument, simulations can help form your intuition; but a purely empirical answer will only receive partial points. How does your answer change for Holm's method? (If a method does not use up its type-1 error budget, it is less powerful than it could have been; it is leaving some error budget on the table.)*

Question 2 (Closed testing with Simes' and Bonferroni tests) *Consider four hypotheses with p-values 0.01, 0.02, 0.1, 0.2 and let's say* $\alpha = 0.1$. *Draw the closed testing graph and indicate the following for each node: (a) [before closure] after running only Simes' test on nodes with an even number of hypotheses and running Bonferroni's test on nodes with an odd number of hypotheses, mark nodes that were rejected; (b) [after closure] after the closed testing procedure, mark the nodes that were rejected. Which nodes would you finally report to have FWER control at level* 0.1*?*

Question 3 (A more conservative procedure than BH) *Consider the index* $\hat{k}_{BH}$ *made by the BH procedure, meaning that it rejects* $P_{(1)}, P_{(2)}, \ldots, P_{(\hat{k}_{BH})}$. *Now consider some other index* $k_0 \leq \hat{k}_{BH}$ *and consider a different procedure that rejects* $P_{(1)}, \ldots, P_{(k_0)}$. *Does this new (more conservative) procedure control FDR? Why or why not? Feel free to use proofs or arguments from class.*

Question 4 (Weighted FWER) *Assume that you assign each of* n *hypotheses a data-independent weight* $w_i > 0$ *such that* $\sum_i w_i = n$. *Prove that the following procedure controls FWER at level* α*: define* $Q_i = P_i/w_i$, *and reject* H_i *if* $Q_i \leq \alpha/n$. *Then, prove that the following procedure also controls FWER at level* α*: sort the* Q_i, *and keep rejecting* $H_{(i)}$ *until for the first time,* $Q_{(i)} > \alpha/\sum_{j=i}^{n} w_{(j)}$. *Is one procedure always more powerful than the other? (Why or why not?)*

Question 5 (Adaptive Bonferroni) *Assume that the null p-values are independent. Pick a* $\lambda \in [\alpha, 1)$ *and define the estimated null proportion as*

$$\widehat{\pi}_0 := \frac{1 + \sum_{i \in [n]} \mathbf{1}(P_i > \lambda)}{n(1 - \lambda)}.$$

Reject H_i *if* $P_i < \alpha/(n\widehat{\pi}_0)$. *Does this procedure control the FWER at level* α*? Does it control PFER at level* α*? When will this procedure be more powerful than Bonferroni? When will this procedure be less powerful than Bonferroni?*