Eric Kong

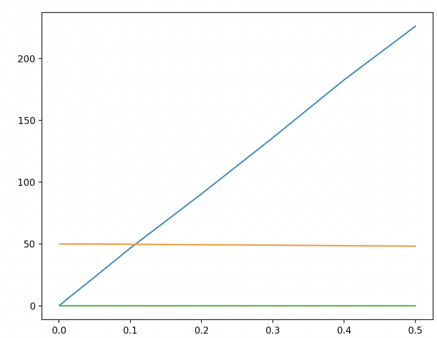
Data 102
September 22, 2019

Homework 2

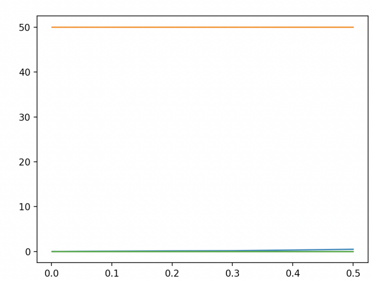
Problem 1

1. Given $\frac{1}{N}\sum_{i=1}^{N} w_i = 1$ then $\sum_{i=1}^{N} w_i = N$ and therefore $w_i < N$.
   Assuming $\alpha \in (0,1)$, the quantity $\alpha\frac{w_i}{N}$ is less than 1.
   Using fact that if P_i is null then, $P(P_i \leq u) = u$ for $u \in [0,1]$.
   So, $u = \frac{\alpha * w_i}{N}$ and $P(P_i \leq u) = \frac{\alpha * w_i}{N}$.
   Since $\frac{w_i}{N} < 1$ then $\alpha\frac{w_i}{N} < \alpha$ and the procedure controls the probability of at least one false discovery under level α.

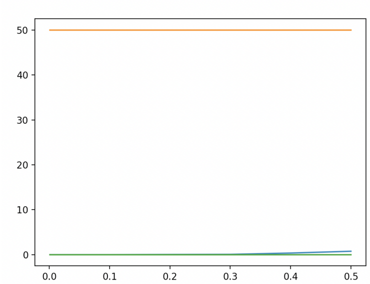
2. $w_i = 1$ implies Bonferroni

3. 

   (a) Uncorrelated Testing

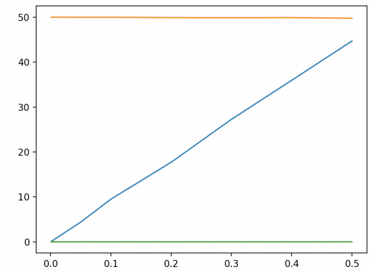
   

   (b) Bonferroni

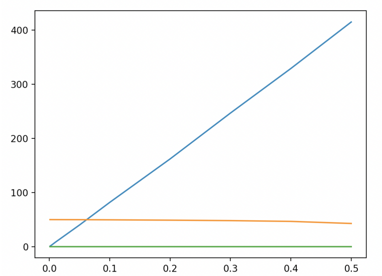
   

   (c) Benjamini-Hochberg

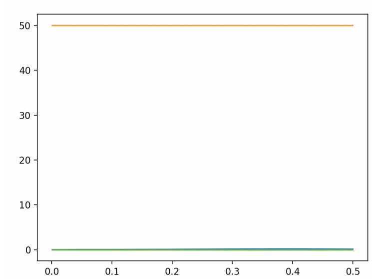
(d) Weighted with $w_i = \frac{2i}{501}$



(e) Weighted with $w_i = \frac{2(501-i)}{501}$



(f) Weighted with $w_i = 0.5$ for $1 \leq i \leq 450$ and $w_i = 5.5$ for $i > 450$



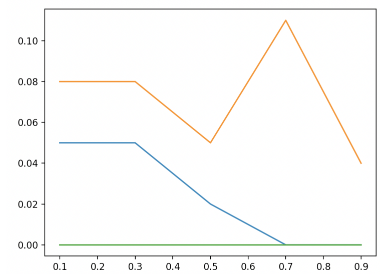
Problem 2

1.

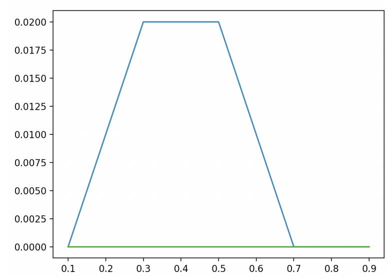
Problem 3

1. Using LORD procedure

   (a) Average FDP over 100 trials at each π

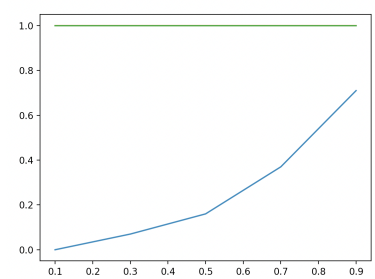
   (b) Average Sensitivity over 100 trials at each π

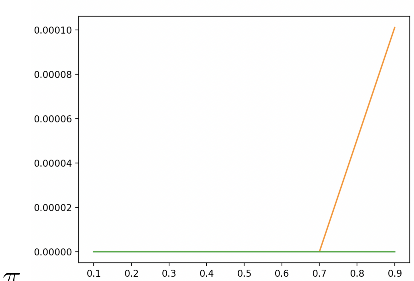
   (c) Highest average sensitivity came from scenario (i).

2. Using Benjamini-Hochberg:

   (a) Average FDP over 100 trials at each π

   (b) Average Sensitivity over 100 trials at each π

   (c) Sensitivity in BH for scenarios (ii) and (iii) are similar to the LORD procedure in that they are very close to 0.

Problem 4

1. Given that all P-values are independent, the case of removing a random P_i from the set to obtain $P^{(-i)}$, results in the events:
$\{P_i \leq \frac{\alpha}{N}r, R = r\}$ and $\{P_i \leq \frac{\alpha}{N}r, R^{(-i)} = r - 1\}$ being equal.
The case breaks down into two parts. Given these two cases, the argument should hold.
   (a) If the randomly removed P-value, P_i, is greater than or less than the maximum sorted P-value such that BH holds, then the set of rejected hypotheses will not change between P and $P^{(-i)}$.
   (b) If the randomly removed P-value, P_i, is equal to the would be maximum, R. In this case the set of rejected hypotheses also remains the same because from the definition of $R^{(-i)}$, the condition that
   $P^{(-i)}_{(j)} \leq \frac{\alpha}{N}(j + 1)$
   accounts for the size difference of 1 (single missing P-value) by requiring the right side of the comparison to be multiplied by the index + 1 - i.e. $(j + 1)$.

2. $\frac{1}{R}\sum_{i\in H_0} 1\{P_i \leq \frac{\alpha}{N}R, R > 0\} = \frac{1}{R}\sum_{i\in H_0} 1\{P_i \leq \frac{\alpha}{N}R\}1\{R > 0\}$

   $= \frac{1}{R}\sum_{i\in H_0} 1\{P_i \leq \frac{\alpha}{N}R\}\sum_{r=1}^{N} 1\{R = r\} = \frac{1}{R}\sum_{i\in H_0}\sum_{r=1}^{N} 1\{P_i \leq \frac{\alpha}{N}R\}1\{R = r\}$

   $= \sum_{i\in H_0}\sum_{r=1}^{N} \frac{1}{r}1\{P_i \leq \frac{\alpha}{N}r, R = r\}$ and finally from part (a)

   $= \sum_{i\in H_0}\sum_{r=1}^{N} \frac{1}{r}1\{P_i \leq \frac{\alpha}{N}r, R^{(-i)} = r - 1\}$

3. Using the fact that the expectation of an indicator is just its probability:
FDR = $\mathbb{E}$(FDP)

   $\mathbb{E}(\text{FDP}) = \mathbb{E}(\sum_{i\in H_0}\sum_{r=1}^{N} \frac{1}{r}1\{P_i \leq \frac{\alpha}{N}r, R^{(-i)} = r - 1\})$

   $= \sum_{i\in H_0}\sum_{r=1}^{N} \frac{1}{r}P(P_i \leq \frac{\alpha}{N}r, R^{(-i)} = r - 1) = \text{FDR}$

4.

5.