STA200C HW4

Reference:

- Generalizations of the Familywise Error Rate by E. L. Lehmann and Joseph P. Romano `https://arxiv.org/pdf/math/0507420.pdf`
- The control of the false discovery rate in multiple testing under dependency by Yoav Benjamini and Daniel Yekutieli `https://projecteuclid.org/euclid.aos/1013699998`

1. Q1

   Fix any P and suppose H_i with $i \in I = I(P)$ are true and the remainder false, with $|I|$ denoting the cardinality of I. Let V be the number of false rejections. Then, by Markov's inequality,

$$\begin{aligned} P(V \geq k) &\leq \frac{\mathrm{E}(V)}{k} \\ &= \frac{\mathrm{E}(\sum_{i \in I(P)} I_{p_i \leq k\alpha/n})}{k} \\ &= \frac{P(p_i \leq k\alpha/n)}{k} \\ &\leq \sum_{i \in I(P)} \frac{k\alpha/n}{k} \\ &= |I(P)|\frac{\alpha}{n} \\ &\leq \alpha. \end{aligned}$$

2. Q2

   Fix any P and let $I(P)$ be the indices of the true null hypothesis. Assume $|I(P)| \geq k$ or there is nothing to prove. Order the p-values corresponding to the $|I(P)|$ true null hypothesis; call them

$$\hat{q}_{(1)} \leq \cdots \leq \hat{q}_{|I(P)|}.$$

   Let j be the smallest (random) index satisfying $\hat{p}_{(j)} = \hat{q}_{(k)}$, so

$$k \leq j \leq n - |I(P)| + k \tag{1}$$

   because the largest possible index j occurs when all the smallest p-values cooresponding to the $n - |I(P)|$ false null hypotheses and the next $|I(P)|$ p-values corerespond to the true null hypotheses. So $\hat{p}_{(j)} = \hat{q}_{(k)}$. Then our modified Holm procedure commits at least k false rejections if and only if

$$\hat{p}_{(1)} \leq \alpha_1, \qquad \hat{p}_{(2)} \leq \alpha_2, \qquad \ldots, \qquad \hat{p}_{(j)} \leq \alpha_j$$

which certainly implies that

$$\hat{q}_{(k)} = \hat{p}_{(j)} \leq \alpha_j = \frac{k\alpha}{n+k-j}$$

But by (1),

$$\frac{k\alpha}{s+k-j} \leq \frac{k\alpha}{|I(P)|}.$$

So the probability of at least k false rejections is bounded above by

$$P\left(\hat{q}_{(k)} \leq \frac{k\alpha}{|I(P)|}\right).$$

By result of Question 1 the chance that the kth largest among $I(P)$ p-values is less than or equal to $k\alpha/|I(P)|$ is less than of equal to α.

3. Q3 Let $U_1, \ldots, U_n$ be i.i.d. uniform random variable on $[0,1]$. Then U_i's satisfy

$$P(U_i \leq \alpha) \leq \alpha.$$

The set of U_i's can be considered as the p-values of n hypotheses where all null hypothesis is true and those p-values are independent. We can apply BH procedure with α level to those p-values. Let R be the number of rejected hypotheses. Then

$$\mathrm{FDP} = \begin{cases} 1 & \text{if } R \geq 1 \\ 0 & \text{if } R = 0 \end{cases}$$

Notice that $P(U_{(i)} \geq i\alpha/n,\ i = 1, \ldots, n) = P(R = 0)$. Therefore

$$\begin{aligned} P(U_{(i)} \geq i\alpha/n,\ i = 1, \ldots, n) &= P(R = 0) \\ &= 1 - P(R \geq 1) \\ &= 1 - \mathrm{E}(\mathrm{FDP}) \\ &= 1 - \mathrm{FDR} \\ &\geq 1 - \alpha. \end{aligned}$$

4. Q4 Let

$$C_r^{(1)} = \left\{ p_{(1)}^{(1)}, \ldots, p_{(r-1)}^{(1)} \leq \frac{qr}{n}, p_{(r)}^{(1)} > \frac{q(r+1)}{n}, \ldots, p_{(n-1)}^{(1)} > q \right\}.$$

Let $0 < i < j \leq n$.

When $C_i^{(1)}$ happens,

$$p_{(1)}^{(1)}, \ldots, p_{(i-1)}^{(1)} \leq \frac{qi}{n}, p_{(i)}^{(1)} > \frac{q(i+1)}{n}, \ldots, p_{(n-1)}^{(1)} > q$$

Then we have $p_{j-1}^{(1)} > \frac{qj}{n}$, but not $p_{j-1}^{(1)} \leq \frac{qj}{n}$, i.e. $C_j^{(1)}$ does not happen.

When $C_j^{(1)}$ happens,

$$p_{(1)}^{(1)}, \ldots, p_{(j-1)}^{(1)} \leq \frac{qj}{n}, p_{(j)}^{(1)} > \frac{q(j+1)}{n}, \ldots, p_{(n-1)}^{(1)} > q$$

Then we have $p_i^{(1)} \leq \frac{q(i+1)}{n}$, but not $p_i^{(1)} > \frac{q(i+1)}{n}$, i.e. $C_i^{(1)}$ does not happen.

Therefore, $C_i^{(1)}$ and $C_j^{(1)}$ are disjoint for any $i < j$.

Let $0 < j \leq n$. Suppose $\mathbf{p} \in C_j^{(1)}$, i.e.,

$$p_{(1)}^{(1)}, \ldots, p_{(j-1)}^{(1)} \leq \frac{qj}{n}, p_{(j)}^{(1)} > \frac{q(j+1)}{n}, \ldots, p_{(n-1)}^{(1)} > q$$

Consider a vector $\mathbf{q}$ such that $q_k \geq p_k$ for all $k = 1, \ldots, n$. We know that $q_{(k)} \geq p_{(k)}$ for all $k = 1, \ldots, n$. Firstly we have

$$q_{(k)}^{(1)} \geq p_{(k)}^{(1)} > \frac{q(k+1)}{n} \quad \text{for all } k = j, \ldots, n.$$

Then we find the smallest i satisfying

$$q_{(k)}^{(1)} > \frac{q(k+1)}{n} \quad \text{for all } k = i, \ldots, j.$$

Such i must exist because j is a candidate. Since i is the smallest index satisfying the above inequality, $i - 1$ would give the below inequalities

$$q_{(1)}^{(1)} \leq q_{(2)}^{(1)} \leq \cdots \leq q_{(i-1)}^{(1)} \leq \frac{qi}{n}$$

Therefore, $\mathbf{q} \in C_i^{(1)}$.

We have shown that: if $\mathbf{p} \in C_j^{(1)}$ and $\mathbf{q} \geq \mathbf{p}$ pointwisely, then $\mathbf{q} \in C_i^{(1)}$ for some $i \leq j$. In other words, if $\mathbf{p} \in C_j^{(1)}$ and $\mathbf{q} \geq \mathbf{p}$, then $\mathbf{q} \in C_1^{(1)} \cup \cdots \cup C_j^{(1)}$.