Procedures Controlling the k-FDR Using Bivariate Distributions of the Null p-Values

Sanat K. Sarkar and Wenge Guo

Temple University and New Jersey Institute of Technology

Supplementary Material

S1 Proofs

PROOF OF LEMMA 2.1. Let $\hat{P}_{(1)} \leq \cdots \leq \hat{P}_{(n_0)}$ be the ordered values of $\{\hat{P}_1, \ldots, \hat{P}_{n_0}\}$. Note that $\hat{P}_{(j)} \leq P_{(n-n_0+j)}$ for all $1 \leq j \leq n_0$. Therefore, the original stepwise procedure rejects less number of null hypotheses, and hence has less number of falsely rejected null hypotheses, than the corresponding stepwise procedure where the p-values corresponding to the false null hypotheses are all very close to zero. Thus, the lemma follows.

■

PROOF OF THEOREM 3.1.

$$
\begin{aligned}
k\text{-FDR} &= E\left\{\frac{V}{R} \cdot I\left(V \geq k\right)\right\} = E\left\{\sum_{r=k}^{n} \frac{1}{r} \sum_{i=1}^{n_0} I\left(\hat{P}_i \leq \alpha_r, V \geq k, R = r\right)\right\} \\
&= \sum_{i=1}^{n_0} \sum_{r=k}^{n} \frac{1}{r} \Pr\left\{\hat{P}_i \leq \alpha_r, V \geq k, R = r\right\} \\
&= \frac{\alpha_k}{k} \sum_{i=1}^{n_0} \sum_{r=k}^{n} P\left\{V \geq k, R = r \mid \hat{P}_i \leq \alpha_r\right\} \; . \qquad \text{(S1.1)}
\end{aligned}
$$

Now, for any $r > k$ and $i \leq n_0$,

$$
\begin{aligned}
& \Pr\left\{V \geq k, R = r \mid \hat{P}_i \leq \alpha_r\right\} \\
= \; & \Pr\left\{V \geq k, R \geq r \mid \hat{P}_i \leq \alpha_r\right\} - \Pr\left\{V \geq k, R \geq r+1 \mid \hat{P}_i \leq \alpha_r\right\} \\
\leq \; & \Pr\left\{V \geq k, R \geq r \mid \hat{P}_i \leq \alpha_{r-1}\right\} - \Pr\left\{V \geq k, R \geq r+1 \mid \hat{P}_i \leq \alpha_r\right\} \; .
\end{aligned}
$$

The inequality follows from the assumption (1), since $\{V \geq k, R \geq r\}$ is a decreasing set for a stepwise procedure. Thus, we get

$$\begin{aligned}
k\text{-FDR} \;\leq\;& \frac{\alpha_k}{k}\sum_{i=1}^{n_0} \Pr\left\{V \geq k, R = k \mid \hat{P}_i \leq \alpha_k\right\} + \\
& \frac{\alpha_k}{k}\sum_{i=1}^{n_0}\sum_{r=k+1}^{n} \Pr\left\{V \geq k, R = r \mid \hat{P}_i \leq \alpha_r\right\} \\
\leq\;& \frac{\alpha_k}{k}\sum_{i=1}^{n_0} \Pr\left\{V \geq k, R = k \mid \hat{P}_i \leq \alpha_k\right\} + \\
& \frac{\alpha_k}{k}\sum_{i=1}^{n_0} \Pr\left\{V \geq k, R \geq k+1 \mid \hat{P}_i \leq \alpha_k\right\} \\
=\;& \frac{\alpha_k}{k}\sum_{i=1}^{n_0} \Pr\left\{V \geq k \mid \hat{P}_i \leq \alpha_k\right\} = \frac{1}{k}\sum_{i=1}^{n_0} \Pr\left\{V \geq k, \hat{P}_i \leq \alpha_k\right\} \;. \qquad \text{(S1.2)}
\end{aligned}$$

Applying Lemma 2.1 to (S1.2), we get

$$k\text{-FDR} \leq \frac{1}{k}\sum_{i=1}^{n_0} P\left\{\hat{R}_{n_0} \geq k, \hat{P}_i \leq \alpha_k\right\}.$$

Let $\hat{R}^{(-i)}_{n_0-1}$ denote the number of rejections in the corresponding stepwise procedure based on the ordered values $\hat{P}^{(-i)}_{(1)} \leq \cdots \leq \hat{P}^{(-i)}_{(n_0-1)}$ of $\left\{\hat{P}_1, \ldots, \hat{P}_{n_0}\right\} \setminus \left\{\hat{P}_i\right\}$ and the $n_0 - 1$ critical values $\alpha_{n-n_0+2} \leq \cdots \leq \alpha_n$. Then, for any $k \leq r \leq n_0$, we notice that, when our procedure is a stepup procedure

$$\begin{aligned}
& \left\{\hat{R}_{n_0} = r, \hat{P}_i \leq \alpha_k\right\} \\
=\;& \left\{\hat{P}_{(r)} \leq \alpha_{n-n_0+r}, \hat{P}_{(r+1)} > \alpha_{n-n_0+r+1}, \ldots, \hat{P}_{(n_0)} > \alpha_n,\ \hat{P}_i \leq \alpha_k\right\} \\
=\;& \left\{\hat{P}^{(-i)}_{(r-1)} \leq \alpha_{n-n_0+r}, \hat{P}^{(-i)}_{(r)} > \alpha_{n-n_0+r+1}, \ldots, \hat{P}^{(-i)}_{(n_0-1)} > \alpha_n,\ \hat{P}_i \leq \alpha_k\right\} \\
=\;& \left\{\hat{R}^{(-i)}_{n_0-1} = r-1, \hat{P}_i \leq \alpha_k\right\}; \qquad \text{(S1.3)}
\end{aligned}$$

whereas, for a stepdown procedure, we have

$$\begin{aligned}
& \left\{\hat{R}_{n_0} = r, \hat{P}_i \leq \alpha_k\right\} \\
=\;& \left\{\hat{P}_{(1)} \leq \alpha_{n-n_0+1}, \cdots, \hat{P}_{(r)} \leq \alpha_{n-n_0+r}, \hat{P}_{(r+1)} > \alpha_{n-n_0+r+1},\ \hat{P}_i \leq \alpha_k\right\} \\
\subseteq\;& \left\{\hat{P}^{(-i)}_{(1)} \leq \alpha_{n-n_0+2}, \cdots, \hat{P}^{(-i)}_{(r-1)} \leq \alpha_{n-n_0+r}, \hat{P}^{(-i)}_{(r)} > \alpha_{n-n_0+r+1},\ \hat{P}_i \leq \alpha_k\right\} \\
=\;& \left\{\hat{R}^{(-i)}_{n_0-1} = r-1, \hat{P}_i \leq \alpha_k\right\}. \qquad \text{(S1.4)}
\end{aligned}$$

Thus, for a stepwise (stepup or stepdown) procedure, we get

$$\begin{aligned} k\text{-FDR} &\le \frac{1}{k}\sum_{i=1}^{n_0}\sum_{r=k}^{n_0} \Pr\left\{\hat{R}^{(-i)}_{n_0-1} = r-1, \hat{P}_i \le \alpha_k\right\} && \text{(S1.5)} \\ &= \frac{1}{k}\sum_{i=1}^{n_0}\sum_{j(\neq i)=1}^{n_0}\sum_{r=k}^{n_0} \frac{1}{r-1}\Pr\left\{\hat{R}^{(-i)}_{n_0-1} = r-1, \hat{P}_i \le \alpha_k, \hat{P}_j \le \alpha_{n-n_0+r}\right\}. \end{aligned}$$

As explained in (S1.3) and (S1.4),

$$\left\{\hat{R}^{(-i)}_{n_0-1} = r-1, \hat{P}_i \le \alpha_k, \hat{P}_j \le \alpha_{n-n_0+r}\right\} = \left\{\hat{R}^{(-i,-j)}_{n_0-2} = r-2, \hat{P}_i \le \alpha_k, \hat{P}_j \le \alpha_{n-n_0+r}\right\},$$

for a stepup procedure, and

$$\left\{\hat{R}^{(-i)}_{n_0-1} = r-1, \hat{P}_i \le \alpha_k, \hat{P}_j \le \alpha_{n-n_0+r}\right\} \subseteq \left\{\hat{R}^{(-i,-j)}_{n_0-2} = r-2, \hat{P}_i \le \alpha_k, \hat{P}_j \le \alpha_{n-n_0+r}\right\},$$

for a stepdown procedure, where $\hat{R}^{(-i,-j)}_{n_0-2}$ is the number of rejections in the corresponding stepwise procedure involving $\left\{\hat{P}_1, \ldots, \hat{P}_{n_0}\right\} \setminus \left\{\hat{P}_i, \hat{P}_j\right\}$ and critical values $\alpha_{n-n_0+3} \le \cdots \le \alpha_n$. Thus,

$$k\text{-FDR} \le \frac{(n-n_0+k)\alpha_k}{k^2(k-1)} \sum_{i=1}^{n_0}\sum_{j(\neq i)=1}^{n_0}\sum_{r=k}^{n_0} \Pr\left\{\hat{R}^{(-i,-j)}_{n_0-2} = r-2, \hat{P}_i \le \alpha_k \mid \hat{P}_j \le \alpha_{n-n_0+r}\right\}.$$

The inequality follows from the fact that

$$\frac{n-n_0+r}{r-1} \le \frac{n-n_0+k}{k-1},$$

for all $k \le r \le n_0$. Since for $r > k$,

$$\begin{aligned} &\Pr\left\{\hat{R}^{(-i,-j)}_{n_0-2} = r-2, \hat{P}_i \le \alpha_k \mid \hat{P}_j \le \alpha_{n-n_0+r}\right\} \\ =\ &\Pr\left\{\hat{R}^{(-i,-j)}_{n_0-2} \ge r-2, \hat{P}_i \le \alpha_k \mid \hat{P}_j \le \alpha_{n-n_0+r}\right\} - \Pr\left\{\hat{R}^{(-i,-j)}_{n_0-2} \ge r-1, \right. \\ &\quad \left. \hat{P}_i \le \alpha_k \mid \hat{P}_j \le \alpha_{n-n_0+r}\right\} \\ \le\ &\Pr\left\{\hat{R}^{(-i,-j)}_{n_0-2} \ge r-2, \hat{P}_i \le \alpha_k \mid \hat{P}_j \le \alpha_{n-n_0+r-1}\right\} - \Pr\left\{\hat{R}^{(-i,-j)}_{n_0-2} \ge r-1, \right. \\ &\quad \left. \hat{P}_i \le \alpha_k \mid \hat{P}_j \le \alpha_{n-n_0+r}\right\}, \end{aligned}$$

with the inequality following due to the property (1) and the fact that $\{\hat{R}^{(-i,-j)}_{n_0-2} \geq r-2, \hat{P}_i \leq \alpha_k\}$ is decreasing, we have

$$\begin{aligned}
&\sum_{r=k}^{n_0} \Pr\left\{\hat{R}^{(-i,-j)}_{n_0-2} = r-2, \hat{P}_i \leq \alpha_k \mid \hat{P}_j \leq \alpha_{n-n_0+r}\right\} \\
\leq\ & \Pr\left\{\hat{R}^{(-i,-j)}_{n_0-2} = k-2, \hat{P}_i \leq \alpha_k \mid \hat{P}_j \leq \alpha_{n-n_0+k}\right\} + \\
& \sum_{r=k+1}^{n_0} \left[\Pr\left\{\hat{R}^{(-i,-j)}_{n_0-2} \geq r-2, \hat{P}_i \leq \alpha_k \mid \hat{P}_j \leq \alpha_{n-n_0+r-1}\right\} - \right. \\
& \left. \Pr\left\{\hat{R}^{(-i,-j)}_{n_0-2} \geq r-1, \hat{P}_i \leq \alpha_k \mid \hat{P}_j \leq \alpha_{n-n_0+r}\right\}\right] \\
=\ & \Pr\left\{\hat{R}^{(-i,-j)}_{n_0-2} = k-2, \hat{P}_i \leq \alpha_k \mid \hat{P}_j \leq \alpha_{n-n_0+k}\right\} + \\
& \Pr\left\{\hat{R}^{(-i,-j)}_{n_0-2} \geq k-1, \hat{P}_i \leq \alpha_k \mid \hat{P}_j \leq \alpha_{n-n_0+k}\right\} \\
=\ & \Pr\left\{\hat{R}^{(-i,-j)}_{n_0-2} \geq k-2, \hat{P}_i \leq \alpha_k \mid \hat{P}_j \leq \alpha_{n-n_0+k}\right\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
& k\text{-FDR} \\
\leq\ & \frac{(n-n_0+k)\alpha_k}{k^2(k-1)} \sum_{i=1}^{n_0} \sum_{j(\neq i)=1}^{n_0} \Pr\left\{\hat{R}^{(-i,-j)}_{n_0-2} \geq k-2, \hat{P}_i \leq \alpha_k \mid \right. \\
& \left. \hat{P}_j \leq \alpha_{n-n_0+k}\right\} \\
\leq\ & \frac{(n-n_0+k)\alpha_k}{k^2(k-1)} \sum_{i=1}^{n_0} \sum_{j(\neq i)=1}^{n_0} \Pr\left\{\hat{P}_i \leq \alpha_k \mid \hat{P}_j \leq \alpha_{n-n_0+k}\right\} \\
=\ & \frac{1}{k(k-1)} \sum_{i=1}^{n_0} \sum_{j(\neq i)=1}^{n_0} \Pr\left\{\hat{P}_i \leq \alpha_k, \hat{P}_j \leq \alpha_{n-n_0+k}\right\}. \qquad \text{(S1.6)}
\end{aligned}$$

The theorem then follows by considering the maximum of the right-hand side in (S1.6) over the set of values of n_0.

■

Proof of (8). Since the p-values are independent and $\alpha_i = \{i \vee k\}\beta/n$, $i = 1, \ldots, n$, we see from the first line in (S1.5) that

$$k\text{-FDR} \leq \frac{n_0\beta}{n} \sum_{r=k}^{n_0} \Pr\left\{\hat{R}_{n_0-1} = r-1\right\} = \frac{n_0\beta}{n} \Pr\left\{\hat{R}_{n_0-1} \geq k-1\right\},$$

the desired inequality.

■

PROOF OF THEOREM 5.1. Define $p_{ijr} = P\left\{\hat{P}_i \in [\alpha_{j-1}, \alpha_j], V \geq k, R = r\right\}$, given a set critical values $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_n$. Then, from (S1.1) we have,

$$
\begin{aligned}
k\text{-FDR} &= \sum_{r=k}^{n}\sum_{i=1}^{n_0}\frac{1}{r}\Pr\left\{\hat{P}_i \leq \alpha_r, V \geq k, R = r\right\} = \sum_{r=k}^{n}\sum_{i=1}^{n_0}\sum_{j=1}^{r}\frac{1}{r}p_{ijr} \\
&= \sum_{i=1}^{n_0}\sum_{j=1}^{n}\sum_{r=k\vee j}^{n}\frac{1}{r}p_{ijr} \leq \sum_{i=1}^{n_0}\sum_{j=1}^{n}\frac{1}{k\vee j}\sum_{r=k\vee j}^{n}p_{ijr} \\
&\leq \sum_{i=1}^{n_0}\sum_{j=1}^{n}\frac{1}{k\vee j}\Pr\left\{\hat{P}_i \in [\alpha_{j-1}, \alpha_j], V \geq k\right\} \\
&\leq \sum_{i=1}^{n_0}\sum_{j=1}^{n}\frac{\alpha_j - \alpha_{j-1}}{k\vee j}\ .
\end{aligned}
$$

Let $\alpha_i = (i \vee k)\alpha_k/k$, for some fixed $0 < \alpha_k < 1$. Then,

$$
k\text{-FDR} \leq n_0\frac{\alpha_k}{k}\left\{1 + \sum_{j=k+1}^{n}\frac{1}{j}\right\}\ . \tag{S1.7}
$$

The theorem then follows by considering $\alpha_k = k\alpha/n\left\{1 + \sum_{j=k+1}^{n}\frac{1}{j}\right\}$ in (S1.7).

■

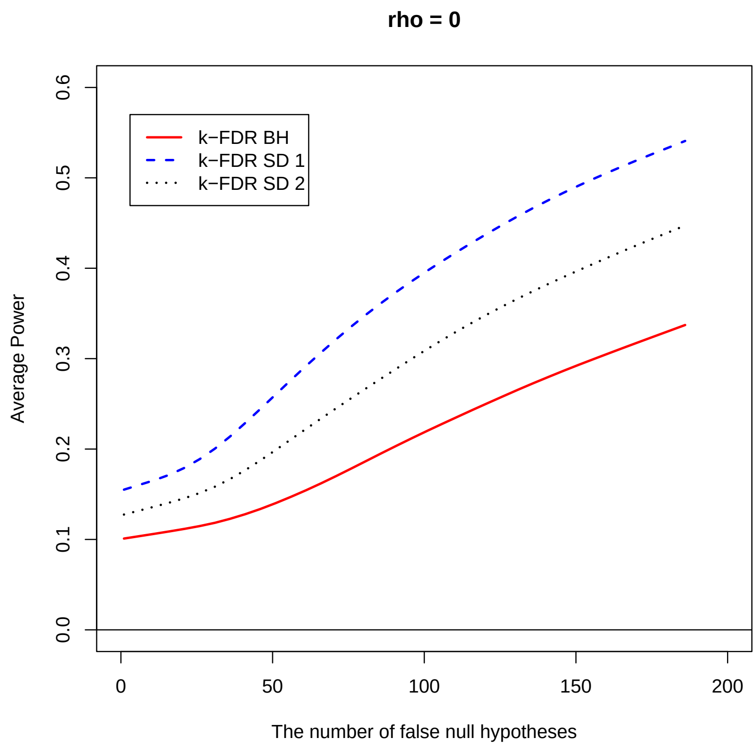

Figure 1: Power of two k-FDR stepdown procedures in the case of independence with parameters $n = 200, k = 5, d = 2$ and $\alpha = 0.05$.