Appendices for "A Bottom-up Approach to Testing Hypotheses That Have a Branching Tree Dependence Structure, with Error Rate Control"

by Yunxiao Li, Yi-Juan Hu, and Glen A. Satten

Appendix

A.1 Weights to account for multiplicity

We first find the mathematical form for weight ω_j (after omitting the level index l from $\omega_{l,j}$) for node j at level l. Recall that the p-value at node j is denoted by p_j. Let $T_1, T_2, \ldots, T_t$ represent all subtrees that contain node j. For example, the node $N_{1,1}$ in Figure 1(a) is contained in subtrees rooted at $N_{2,1}$, $N_{3,1}$ and $N_{4,1}$, which are denoted by T_1, T_2 and T_3, respectively. Let $p_{T_s}^{(-j)}$ be the set of p-values of all level-l nodes, excluding node j, that are contained in subtree T_s. First, each weight ω_j always starts with 1 which counts itself. Then for each subtree T_s, if p_j is the (only) maximum p-value, denoted by the expression $p_j > p_{T_s}^{(-j)}$, then rejecting node j will entail the root node of that subtree to also be rejected and thus add 1 to ω_j. Therefore, we can write ω_j as a sum of indicator functions:

$$\omega_j = 1 + \sum_{s=1}^{t} \mathbb{I}\left(p_j > p_{T_s}^{(-j)}\right). \tag{A1}$$

Note that ω_j depends on p-values at level l and is thus considered to be random (even given the detection events below level l).

Now we prove that the sorted weights at any level l of a complete tree are unique regardless of the ordering of p-values at that level. It is equivalent to prove that, for any h-depth complete tree, the sorted weights at the leaf level are unique; additionally, we can show that the largest weight is h, which is at the node with the largest p-value. We use mathematical induction. First, for any 1-depth tree (i.e., a tree consisting of only one node) the properties hold immediately. Then, assuming the properties hold for any $(h-1)$-depth tree, we consider an h-depth tree. If the root node of the h-depth tree has K children, then deleting the root node would yield K $(h-1)$-level subtrees. Suppose that the leaf node N_j in the h-depth tree has weight ω_j (for a particular ordering of p-values). For a complete tree, each leaf node in the h-depth tree belongs to one of the K $(h-1)$-depth subtrees. Suppose that the node N_j in the $(h-1)$-depth subtree has weight ω'_j. By (A1), we have $\omega_j = \omega'_j + 1$ if N_j has the largest p-value among all leaf nodes of the h-depth tree, and $\omega_j = \omega'_j$ if otherwise. Because the largest p-value in the h-depth tree is also the largest in the $(h-1)$-depth subtree that the p-value belongs to, the corresponding node has weight $(h-1)$ in the $(h-1)$-depth subtree and weight h in the h-depth tree. Thus, the sorted weights at the leaf level of the h-depth tree are obtained from the collection of sorted ω'_js and turning one of the weights $(h-1)$ into h, and must be unique regardless of the ordering of p-values. We thus have proved that the properties hold for any h-depth tree.

A.2 Proof of Theorem 1

Lemma 1 states a property of ω_j, which will be useful in the proof of Theorem 1.

Lemma 1. Suppose node j has weight ω_j as defined in (A1). Assume that node j is under the null hypothesis and so its p-value p_j follows the uniform distribution. Also assume that p_j is independent of all other p-values at level l. Let $\mathcal{B}^{(-j)}$ denote the case that all other p-values excluding p_j belong to a Borel set. For any $\alpha \in (0,1)$, we have

$$\mathbb{E}\left[\omega_j \mathbb{I}\left(\mathcal{B}^{(-j)}, p_j \leq \alpha\right)\right] \leq \frac{\alpha}{1-\alpha} \mathbb{E}\left[\omega_j \mathbb{I}\left(\mathcal{B}^{(-j)}, p_j > \alpha\right)\right].$$

Proof of Lemma 1. Assume that there is no tie among p-values. By the independence assumption of p_j and the other p-values and the uniform distribution of p_j, we have

$\mathbb{E}\left[\mathbb{I}\left(\mathcal{B}^{(-j)}, p_j \le \alpha\right)\right]/\alpha = \Pr\left(\mathcal{B}^{(-j)}\right) = \mathbb{E}\left[\mathbb{I}\left(\mathcal{B}^{(-j)}, p_j > \alpha\right)\right]/(1-\alpha)$. Due to the linearity of ω_j, it then suffices to show for any subtree T_s that

$$\frac{\mathbb{E}\left[\mathbb{I}\left(\mathcal{B}^{(-j)}, p_j > p_{T_s}^{(-j)}, p_j \le \alpha\right)\right]}{\Pr\left(p_j \le \alpha\right)} \le \frac{\mathbb{E}\left[\mathbb{I}\left(\mathcal{B}^{(-j)}, p_j > p_{T_s}^{(-j)}, p_j > \alpha\right)\right]}{\Pr\left(p_j > \alpha\right)}. \tag{A2}$$

By the mean value theorem, there exist p_j^* and p_j^{**}, where $0 < p_j^* < \alpha < p_j^{**} < 1$, such that the left hand side of (A2) becomes

$$\int_0^\alpha \Pr\left(\mathcal{B}^{(-j)}, p_j > p_{T_s}^{(-j)} \mid p_j\right) dF(p_j) \Big/ \int_0^\alpha 1 dF(p_j) = \Pr\left(\mathcal{B}^{(-j)}, p_j^* > p_{T_s}^{(-j)}\right)$$

and the right hand side becomes

$$\int_\alpha^1 \Pr\left(\mathcal{B}^{(-j)}, p_j > p_{T_s}^{(-j)} \mid p_j\right) dF(p_j) \Big/ \int_\alpha^1 1 dF(p_j) = \Pr\left(\mathcal{B}^{(-j)}, p_j^{**} > p_{T_s}^{(-j)}\right).$$

The fact that $p_j^* \le p_j^{**}$ gives (A2). □

Lemma 2 is essentially the summation by parts formula and will also be used in the proof of Theorem 1.

Lemma 2. Suppose $\{a_1, \dots, a_n\}$ and $\{b_1, \dots, b_n\}$ are two sets of real numbers. Then,

$$\sum_{k=1}^n a_k b_k = \sum_{k=1}^n (a_k - a_{k-1}) B_k,$$

where $B_k = \sum_{i=k}^n b_i$ and $a_0 = 0$.

Proof of Theorem 1. This proof is adapted from Gavrilov, Benjamini, and Sarkar (2009) with modifications for our multi-level, bottom-up procedure. The false selection proportion

(FSP) is written as

$$\begin{aligned}\text{FSP} &= \frac{\sum_{l=1}^{L}\sum_{j=1}^{n_l^*}\omega_{l,j}V_{l,j}^m}{\left(\sum_{l=1}^{L}\sum_{j=1}^{n_l^*}\omega_{l,j}R_{l,j}\right)\bigvee 1} \le \sum_{l=1}^{L}\frac{\sum_{j=1}^{n_l^*}\omega_{l,j}V_{l,j}^m}{(\sum_{l'=1}^{l}\sum_{j=1}^{n_{l'}^*}\omega_{l',j}R_{l',j})\bigvee 1} \\ &= \sum_{l=1}^{L}\frac{\sum_{j=1}^{n_l^*}\omega_{l,j}V_{l,j}^m}{(D_{l-1}+\sum_{j=1}^{n_l^*}\omega_{l,j}R_{l,j})\bigvee 1} = \sum_{l=1}^{L}\frac{\sum_{j=1}^{n_l^*}\omega_{l,j}V_{l,j}^m}{D_{l-1}+(\sum_{j=1}^{n_l^*}\omega_{l,j}R_{l,j}\bigvee 1)}.\end{aligned}$$

The key step to prove Theorem 1 is to show that for every level l

$$\mathbb{E}\left[\frac{\sum_{j=1}^{n_l^*}\omega_{l,j}V_{l,j}}{D_{l-1}+(\sum_{j=1}^{n_l^*}\omega_{l,j}R_{l,j}\bigvee 1)}\middle|\mathcal{G}_{l-1}\right] \le q_l, \tag{A3}$$

where $\mathcal{G}_{l-1}$ represents detection events below level l. The inequality (A3) does not guarantee the control of FSR at q_l at each level l because of the cumulative effect of D_{l-1}, which establishes a dependence between the nodes detected at different levels. However, the inequality (A3) leads to the control of overall FSR at q:

$$\text{FSR} = \mathbb{E}(\text{FSP}) \le \sum_{l=1}^{L}\mathbb{E}\left\{\mathbb{E}\left[\frac{\sum_{j=1}^{n_l^*}\omega_{l,j}V_{l,j}}{D_{l-1}+(\sum_{j=1}^{n_l^*}\omega_{l,j}R_{l,j}\bigvee 1)}\middle|\mathcal{G}_{l-1}\right]\right\} \le \sum_{l=1}^{L}q_l = q.$$

To prove (A3) at level l, we omit the level index l for simplicity of exposition. Thus we rewrite the level-l p-values to be $p_1,\ldots,p_{n^*}$, ordered p-values $p_{(1)} \le \cdots \le p_{(n^*)}$, and thresholds defined in (3) $\alpha_1 \le \cdots \le \alpha_{n^*}$. We use D_{-1} for D_{l-1}. We also omit $\mathcal{G}_{l-1}$ by acknowledging that the ensuing arguments are always conditional on the detection events below level l. Further, we denote the set $\{p_{(1)} \le \alpha_1,\ldots,p_{(k)} \le \alpha_k\}$ by $\mathcal{A}_k$ ($k = 1,\ldots,n^*$), which represents the case that the first k ordered p-values are each below the first k thresholds.

We introduce more notation related to the weights $\omega_1 \ldots, \omega_{n^*}$ for level-l nodes. We denote the relative ordering of p-values $p_1,\ldots,p_{n^*}$ by $\mathcal{O}$. The weights defined in (A1) are thus uniquely determined given the detection events at lower levels $\mathcal{G}_{l-1}$ as well as the ordering $\mathcal{O}$. Let $\omega_{[j]}$ be the weight corresponding to the j-th smallest p-value $p_{(j)}$ and $C_k = \sum_{j=1}^{k}\omega_{[j]}$ for

$k = 1, \ldots, n^*$. Thus, C_k is also deterministic given $\{\mathcal{G}_{l-1}, \mathcal{O}\}$. Let $\omega_{(1)} \leq \omega_{(2)} \leq \cdots \leq \omega_{(n^*)}$ denote the sorted weights by their own values; note that $\omega_{(j)}$ is often different from $\omega_{[j]}$. As illustrated in Section 2.2.2, $\omega_{(j)}$ is deterministic given $\mathcal{G}_{l-1}$ under Condition (C1), regardless of the ordering of p-values. Denote $c_k = \sum_{j=1}^{k} \omega_{(j)}$ and $\bar{c}_k = \sum_{j=k}^{n^*} \omega_{(j)}$. Hence, $c_k \leq C_k$, $\bar{c}_k \geq C_{n^*-k+1}$, and the threshold α_k satisfies

$$\frac{\alpha_k}{1-\alpha_k} \leq \frac{D_{-1} + \sum_{j=1}^{k} \omega_{(j)}}{\sum_{j=k}^{n^*} \omega_{(j)}} q_l = \frac{D_{-1} + c_k}{\bar{c}_k} q_l. \tag{A4}$$

Now we prove (A3). The left hand side of (A3) can be rewritten as

$$\sum_{j \in \mathcal{H}_0} \sum_{k=1}^{n^*} \mathbb{E}\left[\frac{\omega_j \mathbb{I}\left(\mathcal{A}_k, p_{(k+1)} > \alpha_{k+1}, p_j \leq \alpha_k\right)}{D_{-1} + C_k}\right],$$

where $\mathcal{H}_0$ denotes the set of modified null hypotheses at level l, and C_k by definition counts the number of all rejections that are produced by rejecting the top k smallest p-values. Replacing the expectation by double expectations that first conditions on $\mathcal{O}$ yields

$$\sum_{j \in \mathcal{H}_0} \sum_{k=1}^{n^*} \mathbb{E}_{\mathcal{O}}\left\{\omega_j \mathbb{E}\left[\frac{\mathbb{I}\left(\mathcal{A}_k, p_{(k+1)} > \alpha_{k+1}, p_j \leq \alpha_k\right)}{D_{-1} + C_k} \middle| \mathcal{O}\right]\right\}, \tag{A5}$$

where C_k becomes a constant given $\mathcal{O}$.

Then, we further replace the inner expectation by double expectations that first conditions on p_j and apply Lemma 2 with $a_k = \mathbb{I}\left(p_j \leq \alpha_k\right)/(D_{-1}+C_k)$, $b_k = \Pr\left(\mathcal{A}_k, p_{(k+1)} > \alpha_{k+1} \middle| p_j, \mathcal{O}\right)$, and $n = n^*$; note that $b_{n^*} = \Pr\left(\mathcal{A}_{n^*} \middle| p_j, \mathcal{O}\right)$. Thus, (A5) becomes

$$\sum_{j \in \mathcal{H}_0} \sum_{k=1}^{n^*} \mathbb{E}_{\mathcal{O}}\left\{\omega_j \mathbb{E}_{p_j}\left[\Pr\left(\mathcal{A}_k \middle| p_j, \mathcal{O}\right) \times \left(\frac{\mathbb{I}\left(p_j \leq \alpha_k\right)}{D_{-1} + C_k} - \frac{\mathbb{I}\left(p_j \leq \alpha_{k-1}\right)}{D_{-1} + C_{k-1}}\right) \middle| \mathcal{O}\right]\right\},$$

which, using $C_k - C_{k-1} = \omega_{[k]}$, can be reorganized as

$$\sum_{j\in\mathcal{H}_0}\sum_{k=1}^{n^*}\mathbb{E}_{\mathcal{O}}\left\{\omega_j\left[\frac{\Pr\left(\mathcal{A}_k, \alpha_{k-1} < p_j \le \alpha_k \middle| \mathcal{O}\right)}{D_{-1}+C_k} - \frac{\omega_{[k]}\Pr\left(\mathcal{A}_k, p_j \le \alpha_{k-1}\middle|\mathcal{O}\right)}{(D_{-1}+C_k)(D_{-1}+C_{k-1})}\right]\right\}. \tag{A6}$$

Let $p_{(1)}^{(-j)} \le \cdots \le p_{(n^*-1)}^{(-j)}$ be the ordered p-values after excluding p_j. We denote the set $\left\{p_{(1)}^{(-j)} \le \alpha_1, \ldots, p_{(k-1)}^{(-j)} \le \alpha_{k-1}\right\}$ by $\mathcal{B}_{k-1}^{(-j)}$, which represents the case that the first $(k-1)$ ordered p-values after excluding p_j are each below the first $(k-1)$ thresholds. We note two facts that relate $\mathcal{A}_k$ and $\mathcal{B}_{k-1}^{(-j)}$. First, $\mathcal{A}_k$ and $\{\alpha_{k-1} < p_j\}$ together imply that p_j cannot be among the first $(k-1)$ smallest p-values and thus the first $(k-1)$ ordered p-values before and after excluding p_j remain the same set. In addition, for any $j \in \{(k), \ldots, (n^*)\}$, $\{p_j \le \alpha_k\}$ implies $\left\{p_{(k)} \le \alpha_k\right\}$. Thus we have $\Pr\left(\mathcal{A}_k, \alpha_{k-1} < p_j \le \alpha_k\middle|\mathcal{O}\right) = \Pr\left(\mathcal{B}_{k-1}^{(-j)}, \alpha_{k-1} < p_j \le \alpha_k\middle|\mathcal{O}\right)$. Second, $\mathcal{B}_{k-1}^{(-j)}$ is a subset of $\mathcal{A}_k$ when $p_j \le \alpha_{k-1}$, which yields $\Pr\left(\mathcal{A}_k, p_j \le \alpha_{k-1}\middle|\mathcal{O}\right) \ge \Pr\left(\mathcal{B}_{k-1}^{(-j)}, p_j \le \alpha_{k-1}\middle|\mathcal{O}\right)$. By the two facts, we see that expression (A6) is less than

$$\sum_{j\in\mathcal{H}_0}\sum_{k=1}^{n^*}\mathbb{E}_{\mathcal{O}}\left\{\omega_j\left[\frac{\Pr\left(\mathcal{B}_{k-1}^{(-j)}, \alpha_{k-1} < p_j \le \alpha_k\middle|\mathcal{O}\right)}{D_{-1}+C_k} - \frac{\omega_{[k]}\Pr\left(\mathcal{B}_{k-1}^{(-j)}, p_j \le \alpha_{k-1}\middle|\mathcal{O}\right)}{(D_{-1}+C_k)(D_{-1}+C_{k-1})}\right]\right\},$$

which can be reorganized as

$$\sum_{j\in\mathcal{H}_0}\sum_{k=1}^{n^*}\mathbb{E}_{\mathcal{O}}\left\{\omega_j\mathbb{E}_{p_j}\left[\Pr\left(\mathcal{B}_{k-1}^{(-j)}\middle|p_j,\mathcal{O}\right)\times\left(\frac{\mathbb{I}\left(p_j\le\alpha_k\right)}{D_{-1}+C_k} - \frac{\mathbb{I}(p_j \le \alpha_{k-1})}{D_{-1}+C_{k-1}}\right)\right]\right\}.$$

After applying Lemma 2 with $a_k = \mathbb{I}\left(p_j \le \alpha_k\right)/(D_{-1}+C_k)$ and $b_k = \Pr\left(\mathcal{B}_{k-1}^{(-j)}, p_{(k)}^{(-j)} > \alpha_k\middle|p_j,\mathcal{O}\right)$, the foregoing expression reduces to

$$\sum_{j\in\mathcal{H}_0}\sum_{k=1}^{n^*}\mathbb{E}_{\mathcal{O}}\left\{\omega_j\mathbb{E}\left[\frac{\mathbb{I}\left(\mathcal{B}_{k-1}^{(-j)}, p_{(k)}^{(-j)} > \alpha_k, p_j \le \alpha_k\right)}{D_{-1}+C_k}\middle|\mathcal{O}\right]\right\}. \tag{A7}$$

Next, we combine the double expectations in (A7) into one and use $c_k \le C_k$ to find that

(A7) is less than

$$\sum_{j\in\mathcal{H}_0}\sum_{k=1}^{n^*}(D_{-1}+c_k)^{-1}\,\mathbb{E}\left[\omega_j\mathbb{I}\left(\mathcal{B}_{k-1}^{(-j)},p_{(k)}^{(-j)}>\alpha_k,p_j\le\alpha_k\right)\right]. \tag{A8}$$

Now the weight ω_j is considered random again. According to Lemma 1 with $\mathcal{B}^{(-j)}=\left\{\mathcal{B}_{k-1}^{(-j)},p_{(k)}^{(-j)}>\alpha_k\right\}$, $\alpha=\alpha_k$, and a null p-value p_j, we obtain

$$\mathbb{E}\left[\omega_j\mathbb{I}\left(\mathcal{B}_{k-1}^{(-j)},p_{(k)}^{(-j)}>\alpha_k,p_j\le\alpha_k\right)\right]\le\frac{\alpha_k}{1-\alpha_k}\mathbb{E}\left[\omega_j\mathbb{I}\left(\mathcal{B}_{k-1}^{(-j)},p_{(k)}^{(-j)}>\alpha_k,p_j>\alpha_k\right)\right].$$

Using (A4) and replacing $\sum_{j\in\mathcal{H}_0}$ by $\sum_{j=1}^{n^*}$, we see (A8) is less than

$$q_l\sum_{k=1}^{n^*}(\bar{c}_k)^{-1}\,\mathbb{E}\left[\sum_{j=1}^{n^*}\omega_j\mathbb{I}\left(\mathcal{B}_{k-1}^{(-j)},p_{(k)}^{(-j)}>\alpha_k,p_j>\alpha_k\right)\right].$$

We see that $\mathcal{B}_{k-1}^{(-j)}$ and $\{p_j>\alpha_k\}$ imply that p_j is not among the top $(k-1)$ smallest p-values. For either $j=(k)$ or $j\in\{(k+1),\ldots,(n^*)\}$, we infer from $p_{(k)}^{(-j)}>\alpha_k$ and $p_j>\alpha_k$ that $p_{(k)}>\alpha_k$. Therefore, $\mathbb{I}\left(\mathcal{B}_{k-1}^{(-j)},p_{(k)}^{(-j)}>\alpha_k,p_j>\alpha_k\right)=\mathbb{I}\left(\mathcal{A}_{k-1},p_{(k)}>\alpha_k\right)$ for $j\in\{(k),\ldots,(n^*)\}$ and 0 for $j\in\{(1),\ldots,(k-1)\}$. Then, the above expression simplifies to

$$q_l\sum_{k=1}^{n^*}(\bar{c}_k)^{-1}\,\mathbb{E}\left[\bar{s}_k\mathbb{I}\left(\mathcal{A}_{k-1},p_{(k)}>\alpha_k\right)\right]\le q_l\sum_{k=1}^{n^*}\Pr\left(\mathcal{A}_{k-1},p_{(k)}>\alpha_k\right)=q_l\left[1-\Pr\left(\mathcal{A}_{n^*}\right)\right]\le q_l.$$

We complete the proof of (A3). □

A.3 Least favorable weights

We calculate the least favorable weights $\widetilde{\boldsymbol{\omega}}_l=(\widetilde{\omega}_{l,(1)},\cdots,\widetilde{\omega}_{l,(n_l^*)})$ for level-l nodes using a recursive algorithm. Recall that a weight $\omega_{l,j}$ counts the number of nodes that are simultaneously rejected if node j is rejected, which includes node j and some of its ancestors; here we allow the ancestors to lie at level $l+1$ and all the way up to the root of the tree to be counted into the weights. For this reason, the sum of the weights at level l is always equal to the number of nodes at and above level l, as rejection of all hypotheses at level l would imply rejection of all hypotheses above level l as well. Our algorithm is built on the following observation: p-value orderings in which a only a few nodes result in rejection of many hypotheses lead to thresholds that are more conservative than p-value orderings in which many nodes each result in rejection of only a few hypotheses.

To develop this intuition into an algorithm, let $\widetilde{\boldsymbol{\omega}}_l^{(h)}$ $(h = 0, 1, 2, \ldots)$ denote the least favorable set of weights against any possible set of weights $\boldsymbol{\omega}^{(l+h)}$, where the subscript (h) indicates that only ancestors up to (including) level $l+h$ are counted. Trivially, $\widetilde{\boldsymbol{\omega}}_l^{(0)} = \boldsymbol{\omega}_l^{(0)} = (1, \ldots, 1)$ with 1 for every node at level l when no ancestors are counted, and this serves as the starting point of the recursive algorithm. At the root node level, denoted by $l + h^*$, $\widetilde{\boldsymbol{\omega}}_l^{(h^*)}$ is the least favorable set of weights $\widetilde{\boldsymbol{\omega}}_l$ that we wish to obtain, and is the end point of the algorithm. We find that $\widetilde{\boldsymbol{\omega}}_l^{(h)}$ can be derived from $\widetilde{\boldsymbol{\omega}}_l^{(h-1)}$ by traversing over every node at level $l + h$, locating the weights in $\widetilde{\boldsymbol{\omega}}_l^{(h-1)}$ that correspond to descendent nodes of that node at level $l+h$, and adding 1 to the weight having the largest numerical value among these descendants; if there exist multiple largest elements, randomly pick one and add 1. For example, to calculate least favorable weights for the bottom level of the tree in Figure 1(b), we obtain $\widetilde{\boldsymbol{\omega}}_l^{(0)} = (1, 1, 1, 1, 1, 1)$, $\widetilde{\boldsymbol{\omega}}_l^{(1)} = (1, 2, 1, 1, 1, 2)$, and $\widetilde{\boldsymbol{\omega}}_l^{(2)} = (1, 2, 1, 1, 1, 3)$, where the elements are unsorted and correspond to $(N_{1,1}, N_{1,2}, \ldots, N_{1,6})$, and finally we obtain the sorted weights $\widetilde{\boldsymbol{\omega}}_l = (1, 1, 1, 1, 2, 3)$. Note that the sum of weights $\widetilde{\boldsymbol{\omega}}_l^{(h)}$ is always larger than the sum of weights $\widetilde{\boldsymbol{\omega}}_l^{(h-1)}$ by the number of nodes at level $l + h$. By incrementing the *largest* weight among descendents of each node, this procedure guarantees that every $\widetilde{\boldsymbol{\omega}}_l^{(h)}$ $(h = 0, 1, \ldots, h^*)$ is the least favorable against any arbitrary $\boldsymbol{\omega}_l^{(h)}$. Although there is a random element in our algorithm when two or more nodes share the largest weight, the *sorted* weights $\widetilde{\boldsymbol{\omega}}_l$ are unique.

A.4 Proof of Theorem 2

Proof of Theorem 2. Let $\widetilde{\omega}_{(1)} \leq \widetilde{\omega}_{(2)} \leq \cdots \leq \widetilde{\omega}_{(n^*)}$ denote sorted least favorable weights after omitting the level index l. Denote $\widetilde{c}_k = \sum_{j=1}^{k} \widetilde{\omega}_{(j)}$ and $\overline{\widetilde{c}}_k = \sum_{j=k}^{n^*} \widetilde{\omega}_{(j)}$. The same arguments in the proof of Theorem 1 can be used except that we use $\widetilde{c}_k$ in place of c_k, $\overline{\widetilde{c}}_k$ in place of $\overline{c}_k$, and the thresholds (5) in place of thresholds (3). □

A.5 An algorithm for separate FSR control

Here, we illustrate our proposal by using the microbiome example and assuming an incomplete phylogenetic tree. We wish to first detect OTUs (level-1 nodes) while controlling the FSR at some rate q_1; then we wish to detect taxa (starting at the species level and continuing up the phylogenetic tree) while controlling the FSR at some rate q_{-1}. To accomplish this, we propose a two-stage procedure. At stage 1, we perform the step-down test for OTUs with thresholds $\{\alpha_{1,j}, j = 1, 2, \ldots, n_1\}$ that satisfy

$$\frac{\alpha_{1,j}}{1-\alpha_{1,j}} = \left(\frac{\sum_{k=1}^{j} \widetilde{\omega}_{1,(k)}}{\sum_{k=j}^{n_1} \widetilde{\omega}_{1,(k)}} \times q_1 \right) \bigwedge \frac{\tau_0}{1-\tau_0}.$$

Note that these are the same thresholds for level 1 as the (one-stage) procedure described in Section 2.3, if the same value of q_1 is used. The use of weights $\{\widetilde{\omega}_{1,(k)}\}$ to account for multiplicity allows us to include in our list of detected OTUs all taxa that are automatically detected after testing OTUs. Thus, the FSR for the OTU level is written as

$$\mathrm{FSR}_{\mathrm{otu}} = \mathbb{E}\left[\frac{\sum_{j=1}^{n_1} \omega_{1,j} V_{1,j}^{m}}{\left(\sum_{j=1}^{n_1} \omega_{1,j} R_{1,j} \right) \bigvee 1} \right].$$

At stage 2, we then apply the one-stage procedure proposed in Section 2.3 to the tree obtained by removing all OTUs as well as those taxa that were detected at stage 1. In this tree, undetected nodes at level 2 are now the leaves, and the p-values for these new leaves are calculated by aggregating the p-values from the undetected OTUs. We use the thresholds

$\{\alpha_{l,j}, l = 2, \ldots, L, j = 1, \ldots, n_l^*\}$ that satisfy

$$\frac{\alpha_{l,j}}{1-\alpha_{l,j}} = \left(\frac{D_{l-1}^{\dagger} + \sum_{k=1}^{j} \widetilde{\omega}_{l,(k)}}{\sum_{k=j}^{n_l^*} \widetilde{\omega}_{l,(k)}} \times q_l \right) \bigwedge \frac{\tau_0}{1-\tau_0},$$

where $D_{l-1}^{\dagger} = \sum_{l'=2}^{l-1} \sum_{j=1}^{n_{l'}^*} \omega_{l',j} R_{l',j}$ for $l \geq 3$ and $D_{l-1}^{\dagger} = 0$ for $l = 2$. The $D_{l-1}^{\dagger}$ differs from D_{l-1} in that $D_{l-1}^{\dagger}$ counts the detected nodes starting from the 2nd level. Using $D_{l-1}^{\dagger}$ in place of D_{l-1} cuts the dependence between level 1 and the remaining levels and also makes the thresholds for $l \geq 2$ more stringent if q_ls stay the same as those in the one-stage procedure. Thus, the FSR we wish to control at the taxa levels is

$$\mathrm{FSR}_{\mathrm{taxa}} = \mathbb{E}\left[\frac{\sum_{l=2}^{L} \sum_{j=1}^{n_l^*} \omega_{l,j} V_{l,j}^m}{\left(\sum_{l=2}^{L} \sum_{j=1}^{n_l^*} \omega_{l,j} R_{l,j} \right) \bigvee 1} \right].$$

Theorem 3 states that this two-stage procedure serves our purpose.

Theorem 3. Under Conditions (C2) and (C3) in Theorem 1, the above two-stage procedure ensures that $\mathrm{FSR}_{\mathrm{otu}} \leq q_1$ and $\mathrm{FSR}_{\mathrm{taxa}} \leq q_{-1} = \sum_{l=2}^{L} q_l$.

Proof of Theorem 3. The OTU-level testing in the two-stage procedure is exactly the same as the OTU-level testing in the one-stage procedure, so we immediately have $\mathrm{FSR}_{\mathrm{otu}} \leq q_1$. Then we rewrite the FSP among all taxa levels as

$$\mathrm{FSP}_{\mathrm{taxa}} = \frac{\sum_{l=2}^{L} V_l}{(\sum_{l=2}^{L} R_l) \bigvee 1} \leq \sum_{l=2}^{L} \frac{V_l}{(\sum_{l'=2}^{l} R_{l'}) \bigvee 1} = \sum_{l=2}^{L} \frac{V_l}{(D_{l-1}^{\dagger} + R_l) \bigvee 1} = \sum_{l=2}^{L} \frac{V_l}{D_{l-1}^{\dagger} + (R_l \bigvee 1)}.$$

We note that $D_{l-1}^{\dagger}$ is deterministic conditioning on the detection events at lower levels, denoted by $\mathcal{G}_{l-1}^{\dagger}$, which excludes the OTU level. Then we follow the same steps in the proof of Theorem 2, replacing D_{l-1} by $D_{l-1}^{\dagger}$ and $\mathcal{G}_{l-1}$ by $\mathcal{G}_{l-1}^{\dagger}$ to obtain $\mathbb{E}\left[V_l / \left\{ D_{l-1}^{\dagger} + (R_l \bigvee 1) \right\} \middle| \mathcal{G}_{l-1}^{\dagger} \right] \leq$

q_l for $l = 2, \ldots, L$. Finally,

$$\mathrm{FSR}_{\mathrm{taxa}} = \mathbb{E}\left(\mathrm{FSP}_{\mathrm{taxa}}\right) \leq \sum_{l=2}^{L} \mathbb{E}\left\{\mathbb{E}\left[\frac{V_l}{D_{l-1}^{\dagger} + (R_l \bigvee 1)}\middle|\mathcal{G}_{l-1}^{\dagger}\right]\right\} \leq \sum_{l=2}^{L} q_l = q_{-1},$$

which implies that FSR among all taxa levels are controlled by q_{-1}. Indeed, this is the same as applying the one-stage testing at FSR q_{-1} to the subtree after removing the whole OTU level and the higher-level taxa that are detected because all of their corresponding OTUs are detected. $\square$

Note that the choice of $(q_1, q_2, \ldots, q_L)$ is at the user's discretion and not necessary to match those in the one-stage procedure. For example, we can set $q_1 = q_{-1} = 5\%$ and choose $q_l = q_{-1} n_l / \left(\sum_{l'=2}^{L} n_{l'}\right)$ for $l = 2, \ldots, L$.