Tree-based Multiple Hypothesis Testing with General FWER Control

David Kim

May 26, 2025

1. Notations

We give online version of notations of Goeman and Solari (2010). First, for each time point i, which is indexed by positive integers and zero, define M_i as statistical model used in ith time points, and $H_{i,\text{new}}$ is a set of null hypotheses which is observed new in the ith time point. $H_{i,\text{new}} = F_{i,\text{new}}(M_i) \cup T_{i,\text{new}}(M_i)$, where $T_{i,\text{new}}(M_i)$ is a collection of true nulls in $H_{i,\text{new}}$ under model M_i, and $F_{i,\text{new}}(M_i) = H_{i,\text{new}} \cap T^c_{i,\text{new}}(M_i)$. We define

$$H_i := \bigcup_{k \leq i} H_{k,\text{new}}$$
$$F_i := \bigcup_{k \leq i} F_{k,\text{new}}$$
$$T_i := \bigcup_{k \leq i} T_{k,\text{new}}.$$

We say $\mathbb{F}_i = \sigma(R_1, ..., R_i)$, where R_i is a random vector which indicates whether or not the nulls on ith time point is rejected or not. For all other notation, we follow them from Goeman and Solari (2010), like R_i as a set of rejected hypotheses by ith time point and N as a successor function, so a function determines the next set of rejected hypotheses. Particularly, we let

$$\lim_{n \to \infty} M_n = M.$$

2. Theoretical Foundations

We first give a sequential rejection principle in our setting, which is an online version of Theorem 1 on Goeman and Solari (2010).

Theorem 1. For every $R_i \subseteq S_i \supset H_i$ for any i, almost surely,

$$N(R_i) \subseteq N(S_i) \cup S_i, \tag{1}$$

and for every M_i, we have

$$P_{M_{i+1}}\Big(N(F_i) \subseteq F_{i+1} \Big| \mathbb{F}_i\Big) \geq 1 - \alpha. \tag{2}$$

Then, for every M,

$$P_M(R_\infty \subseteq F_\infty | \mathbb{F}_\infty) \geq 1 - \alpha.$$

Proof. Proof will be almost same as Theorem 1 of Goeman and Solari (2010), and the only difference will come from using conditional dominated convergence theorem than the original dominate convergence theorem. □

Now, for the tree-based multiple hypotheses testing, we assume logical relationship between the nulls, so the parent null is always an intersection between its children nulls. We also assume independency between the observations on same level, different nodes.

We make clusters for the nulls having the same parent null on the tree. As an online testing procedure, without loss of generality, for the same level of the tree we consider the cluster of nulls from left to right. If we test all of the nulls in each level, then we proceed to the leftest cluster of nulls on the next level.

We consider using adjusted critical value function from Hommel's or Simes', and apply these adjustments locally by cluster to cluster. To show that the online sequential rejection principle holds for our case, we need to show (1)(which is often called the **monotonicity condition**, and (2)(which is often called the **single-step condition**). Since we assumed a logical relationship between the nulls, (1) is already satisfied. Then it is enough to check (2), so check about

$$P_{M_{i+1}}\Big(\bigcup_{H\in T_{i+1}}\{p_H \leq \alpha_H(F_i)\}|\mathbb{F}_i\Big),$$

where p_H is a p-value on the node H, and $\alpha_H(R_i)$ is a critical value function on node H when the collection of rejected hypotheses on ith time point, namely, R_i is given.

When we apply Hommel's adjustment, we will actually use $\alpha_i(R) = \alpha/i$ for some fixed α value, where $\alpha_i(R)$ is an ith smallest critical value(***need to be elaborated further***). Then, since we assumed independency among observations in the same cluster, Simes's inequality holds, so by following arguments of Section 5.3 on Goeman and Solari (2010) will directly imply the desired FWER control on the procedure.

REFERENCES

Jell J. Goeman and Aldo Solari. The sequential rejection principle of familywise error control. *The Annals of Statistics*, 38:6:3782–3810, 2010.