STAT 689: Large-scale and high-dimensional statistical inference **Fall 2024**

Lecture 5

1 Graphical procedures

Bonferroni, Holm's, and Hochberg's procedures are all symmetric procedures. This means that if we change the p-values between the hypotheses, then the rejected hypotheses will also change accordingly. This approach is useful when we have no prior knowledge about the hypotheses being tested, like in a genome-wide association study (GWAS) where we do not know which genes are most likely to be associated with a trait.

There are situations where the hypotheses are asymmetric, and we want to incorporate this asymmetry into our procedures. A typical example is in clinical trials, which often have multiple stages. A significant result in the first stage is typically more important than in later stages. It is usually necessary for FDA approval, regardless of the significance and effect sizes of later stages. Another source of asymmetry may come from prior knowledge about effect sizes. If one alternative is expected to have a large effect size, then we may prioritize testing that hypothesis. Graphical procedures are a family of procedures that use various types of prior knowledge to distribute the "α-budget" of the procedure unevenly and adaptively across the n hypotheses.

We provide two examples before giving a definition of the general graphical procedures.

1.1 Sequential testing

Suppose we can order the hypotheses in a way such that H_1 is the "most promising" hypothesis, the one that is mostly likely to be non-null, and H_2 is the second "most promising" hypothesis. In this case, we can perform the following sequential test

- Step 1. For $j = 1$, compare p_j with α.
- Step 2. If $p_j \leq \alpha$, set $j = j + 1$ and move to step 1. Otherwise, terminate and reject $H_1, \ldots, H_{j-1}$.

We next show that this procedure controls the FWER strongly, which means that the FWER is controlled under all possible arrangements of null and alternative hypotheses. This differs from weak control of the FWER, where the FWER is only controlled under the global null.

Theorem. The sequential test controls the FWER strongly.

Proof. Let H_j be the first true null being rejected (which means H_j is the first true null on the ordered list). Then, we have

$$\begin{aligned}\text{FWER} \leq& \mathbb{P}(\text{reject } H_j)\\ \leq& \mathbb{P}(p_1 \leq \alpha, \ldots, p_{j-1} \leq \alpha, p_j \leq \alpha)\\ \leq& \mathbb{P}(p_j \leq \alpha) = \alpha.\end{aligned}$$

It important that the order of the hypotheses was chosen in advance and independently of the p-values.

1.2 Fallback procedures

Consider the above sequential setting with $\alpha = 0.05$ and

$$p_1 = 0.01, \quad p_2 = 0.02, \quad p_3 = 0.09, \quad p_4 = 0.01.$$

Here H_1, H_2, H_4 are non-null and H_3 is null. As $p_3 > 0.05$, we fail to reject H_4. To overcome this issue, we consider a procedure that considers each hypothesis.

Like the sequential test, suppose we are given an ordered list of hypothesis, say $H_1, \dots, H_n$, where the order is independent of the p-values. Unlike fixed sequence testing, a set of thresholds $\alpha_1, \dots, \alpha_n$ such that $\sum_{i=1}^n \alpha_i = \alpha$ must also be fixed in advance. Once the p-values have been calculated, the fallback procedure does the following:

- We first compare p_1 with α_1. If $p_1 \le \alpha_1$, then we reject H_1 and set α_2 to $\alpha_1 + \alpha_2$. Otherwise, we leave α_2 unchanged.
- We compare p_2 with α_2. If $p_2 \le \alpha_2$, then we reject H_2 and set α_3 to $\alpha_2 + \alpha_3$. Otherwise, we leave α_3 unchanged.
- In general, suppose we have tested $H_1, \dots, H_{j-1}$ and we want to examine H_j. If $p_j \le \alpha_j$, then we reject H_j and set α_{j+1} to $\alpha_j + \alpha_{j+1}$. Otherwise, we leave α_{j+1} unchanged.

Note that the sequential test is a fallback procedure with $\alpha_1 = \alpha$ and $\alpha_i = 0$ for $i > 1$. We next show that the fallback procedure controls the FWER strongly.

Theorem. The fallback procedure controls the FWER strongly.

Proof. Let $1 \le j_1 < j_2 < \cdots < j_{n_0} \le n$ be the indices of the true nulls. We note that

$$\begin{aligned}
\text{FWER} =& \mathbb{P}\left(H_{j_k} \text{ is rejected for some } k = 1, 2, \dots, n_0\right) \\
=& \mathbb{P}\left(H_{j_1} \text{ is rejected}\right) + P\left(H_{j_1} \text{ is not rejected and } H_{j_2} \text{ is rejected}\right) \\
& + \cdots + P\left(H_{j_k} \text{ is not rejected for all } k < n_0 \text{ and } H_{j_{n_0}} \text{ is rejected}\right) \\
\le& \mathbb{P}(p_{j_1} \le \alpha_1 + \cdots + \alpha_{j_1}) + \mathbb{P}(p_2 \le \alpha_{j_1+1} + \cdots \alpha_{j_2}) \\
& + \cdots + \mathbb{P}(p_{j_{n_0}} \le \alpha_{j_{n_0-1}+1} + \cdots + \alpha_{j_{n_0}}) \\
=& \sum_{i=1}^n \alpha_i = \alpha.
\end{aligned}$$

We have a couple of remarks about the above proof.

- As with the sequential test, the important assumption was that the order of hypotheses and the thresholds α_i's were specified independently of the p-values.
- The proof makes no assumptions about the dependencies between the p-values.
- The proof is similar to the proof that closed procedures control the FWER. In both proofs, we consider the subset of $\{1, 2, \dots, n\}$ containing the true nulls.

1.3 General graphical procedures

To specify a graphical procedure, the following must be given

- an error level $\alpha \in (0, 1)$;
- a set of thresholds $\alpha_1, \dots, \alpha_n$ such that $\alpha_i \ge 0$ and $\sum_{i=1}^n \alpha_i \le \alpha$;
- A weighted directed graph $(w_{ij})_{i,j=1}^n$ with $w_{ij} \in [0, 1]$, $w_{ii} = 0$ and $\sum_{j=1}^n w_{ij} \le 1$ for all $i, j = 1, 2, \dots, n$.

We can associate each vertice with a hypothesis H_i. The edge between H_i and H_j has a weight w_{ij}. When $w_{ij} = 0$, we interpret this as an absence of an edge between H_i and H_j. Thus, the condition w_{ii} means we do not allow self-loop. Cycles of length greater than one are allowed.

Given the above inputs, the graphical procedure is an iterative procedure that does the following. At each time step, we have a set of current hypotheses $\{H_i : i \in I\}$. This collection is the set of hypotheses we are yet to reject. First, we check if there exists an index $i \in I$ with $p_i \le \alpha_i$. If no such index exists, then the procedure terminates. If such an $i \in I$ does exist, then we reject H_i and perform the following update to our set of hypotheses, thresholds, and

$$
\begin{aligned}
&I \longleftarrow I \setminus \{i\}, \\
&\alpha_j \longleftarrow \alpha_j + \alpha_i w_{ij} \\
&w_{jk} \longleftarrow \begin{cases} \frac{w_{jk}+w_{ji}w_{ik}}{1-w_{ji}w_{ij}} & \text{if } j \neq k, \\ 0 & \text{if } j = k. \end{cases}
\end{aligned}
$$

We then repeat this procedure with the new set of hypotheses.

When we reject H_i, the budget α_i is transferred to the other hypotheses based on the weights w_{ij}. The updated weights can be given the following interpretation. The previous w_{ij} can be roughly thought of as the transition probabilities for a random walk on $\{H_j : j \in I\}$. Once we reject H_i, we construct a new random walk on $\{H_j : j \in I \setminus \{i\}\}$. In the new random walk, we can either transition from H_j to H_k directly or transition via the deleted state H_i. This gives the numerator $w_{jk} + w_{ji}w_{ik}$. The denominator comes because we do not allow for self-loop. Thus, we must correct for the possibility of a transition from H_j back to H_j via H_i. Many of the procedures we have seen so far fall under the umbrella of graphical procedures.

The Bonferroni method is a graphical procedure with $\alpha_i = \alpha/n$ and $w_{ij} = 0$ for all i and j.

Exercise 5.1: Show that the Holm's procedure is also a graphical procedure with $\alpha_i = \alpha/n$ and $w_{ij} = 1/(n-1)$ for all i and j with $i \neq j$.

The sequential procedure is a graphical procedure with $\alpha_1 = \alpha$, $\alpha_i = 0$ for $i = 2, \ldots, n$ and $w_{i,i+1} = 1$ for $i = 1, \ldots, n-1$ and $w_{ij} = 0$ for all other i, j. Likewise, the fallback procedure is a graphical procedure with the same weights as the fixed sequence procedure but with arbitrary thresholds.

Theorem. Given any thresholds $\alpha_1, \ldots, \alpha_n$ and weights $\{w_{ij}\}_{i,j=1}^n$, the corresponding graphical procedure described above controls the FWER strongly under arbitrary dependencies between the p-values. Furthermore, the set of rejected hypotheses $\mathcal{R}$ does not depend on the order in which the hypotheses are rejected.

Proof. The full proof of this theorem is in Appendix A of Bretz et al. (2009, *Statistics in Medicine*). Theorem 3 follows by showing that every graphical procedure is the closure of a weighted Bonferroni procedure.

One limitation of graphical procedures is their lack of flexibility. A choice of α_i's and w_{ij}'s that works well in one situation may not work well in another. It is still an unresolved issue how to select the best thresholds and weights in a given scenario.

1.4 Weighted Bonferroni procedures and consonance

This section shows that graphical procedures are the closures of weighted Bonferroni tests. The FWER control theorem above will thus be a consequence of this connection. To this end, let us first define the weighted Bonferroni procedure.

Definition. Given $\alpha_1, \ldots, \alpha_n \geq 0$ with $\sum_{i=1}^n \alpha_i \leq \alpha$, the weighted Bonferroni test rejects the global null if $p_i \leq \alpha_i$ for some i.

Using the union bound, it is straightforward to show that the weighted Bonferroni method controls the Type I error at level α for testing the global null. Bonferroni's test is the special case when $\alpha_i = \alpha/n$ for every i.

Since we are interested in closing such tests, we need to have thresholds $\alpha_i(I)$ for every intersection hypothesis $H_I = \cap_{i \in I} H_i$. Thus, suppose that we have thresholds $\alpha_i(I)$ for $i \in I$ and every $I \subseteq \{1, 2, \ldots, n\}$ such that $\sum_{i \in I} \alpha_i(I) \leq \alpha$. Then, based on the weighted Bonferroni, we reject H_I if there exists an $i \in I$ such that $p_i \leq \alpha_i(I)$. In other words,

$$\psi_I = \max \{\mathbf{I}\{p_i \leq \alpha_i(I)\} : i \in I\}.$$

We impose a relationship between $\alpha_i(I)$ and $\alpha_i(J)$ for $i \in J \subseteq I$. If $i \in J \subseteq I$, then p_i is a p-value that will be used in testing H_I and H_J. The set I is greater than the set J. This means that when we test H_I, we have a larger multiple testing problem than when we test H_J. Thus, to control the probability of falsely rejecting a null hypothesis, the threshold $\alpha_i(I)$ should be smaller than $\alpha_i(J)$. This property is called monotonicity and can be written as

$$\alpha_i(I) \leq \alpha_i(J), \quad i \in J \subseteq I.$$

Now suppose the thresholds $\alpha_i(I)$ with $I \subseteq \{1, 2, \ldots, n\}$ and $i \in I$ satisfy the monotonicity. Then for all non-empty I, if H_I is rejected, then there exists $i \in I$ such that for all J such that $i \in J \subseteq I$, H_J is also rejected. To see this, note that if H_I is rejected, there exists an $i \in I$ such that $p_i \leq \alpha_i(I)$. As $\alpha_i(I) \leq \alpha_i(J)$, we must have $p_i \leq \alpha_i(J)$ and hence H_J is rejected.

The closure of the weighted Bonferroni tests satisfies a property named consonance.

Definition. Let $\{\psi_I\}$ be a family of tests for testing $\{H_I : I \subseteq \{1, 2, \ldots, n\}\}$. The family $\{\psi_I\}$ is consonant if for all $I \subseteq \{1, 2, \ldots, n\}$ if $\psi_I = 1$, then there exists $i \in I$ such that for all J with $i \in J \subseteq I$, $\psi_J = 1$.

Exercise 5.2: Show that closing Bonferroni is always consonant.

Without consonance, one could be in a situation where the global null has been rejected, but none of the individual nulls have been rejected.

Example. Suppose that for $i = 1, 2$, we have $X_i \sim N(\theta_i, 1)$ independently. Let $H_i : \theta_i = 0$ for $i = 1, 2$. We know that

$$\begin{aligned} &X_1^2 + X_2^2 \sim \chi_2^2 \text{ under global null,} \\ &X_i^2 \sim \chi_1^2 \text{ under } H_i. \end{aligned}$$

Based on the above observations, we propose to reject the global null when $X_1^2 + X_2^2 > \chi_2^2(1-\alpha)$ and H_i when $X_i^2 > \chi_1^2(1-\alpha)$. We could then close the above tests to obtain a procedure that controls the FWER at α. However, the result tests will be nonconsonant as $2\chi_1^2(1-\alpha) > \chi_2^2(1-\alpha)$ (e.g., for $\alpha = 0.05$, $\chi_1^2(0.95) = 3.84$ and $\chi_2^2(0.95) = 5.99$). Thus, one can construct a scenario where $X_1^2, X_2^2 < \chi_1^2(1-\alpha)$ and $X_1^2 + X_2^2 \geq \chi_2^2(1-\alpha)$. This would lead to a scenario where the global null is rejected, but neither H_1 nor H_2 is rejected.

One solution to ensure consonance is to consider the following procedure:

$$\begin{aligned} &\text{reject } H_{12} \text{ if } \max\{X_1^2, X_2^2\} > m(1-\alpha), \\ &\text{reject } H_i \text{ if } X_i^2 > \chi_1^2(1-\alpha), \end{aligned}$$

where $m(1-\alpha)$ is the $1-\alpha$ quantile of the maximum of two independent χ_1^2 random variables. As $m(1-\alpha) > \chi_1^2(1-\alpha)$, if the global null is rejected, H_1 or H_2 has to be rejected, and hence the procedure is consonant.

The algorithm below for the closed weight Bonferroni avoids the exponential complexity of the closure principle by only computing the thresholds $\alpha_i(I)$ for a small number of subsets I.

1.5 Graphical procedures as weighted Bonferroni procedures

We now show that graphical procedures can be described as the closure of a weighted Bonferroni procedure. Suppose we are given $\{\alpha_i\}$ and $\{w_{ij}\}$. We will define the local thresholds $\alpha_i(I)$ and local weights $w_{ij}(I)$ by backward induction on the size of I. When $J = \{1, 2, \ldots, n\}$, we simply set $\alpha_i(J) = \alpha_i$ and $w_{ij}(J) = w_{ij}$. Now consider $I = J \setminus \{i\}$ for $i \in J$. For any $j, k \in I$, we define

$$\begin{aligned} \alpha_j(I) &= \alpha_j(J) + \alpha_i(J) w_{ij}(J), \\ w_{jk}(I) &= \begin{cases} \frac{w_{jk}(J) + w_{ji}(J) w_{ik}(J)}{1 - w_{ji}(J) w_{ij}(J)} & \text{if } j \neq k, \\ 0 & \text{if } j = k. \end{cases} \end{aligned}$$

In Appendix A of Appendix A of Bretz et al. (2009, *Statistics in Medicine*), it is shown that the above definition does not depend on the order in which indices are removed from $\{1, 2, \ldots, n\}$ to produce I. By the

Algorithm 1 Closed weighted Bonferroni procedure

Require: A function that can compute $\alpha_i(I)$ for any $I \subseteq [n]$ and $i \in I$, which satisfies the monotonicity.

```
R ⟵ ∅
I ⟵ {1, 2, . . . , n}
for j ∈ I do
    Compute α_j(I)
end for
while There exists i ∈ I with p_i ≤ α_i(I) do
    Find i such that p_i ≤ α_i(I)
    R ⟵ R ∪ {i}
    I ⟵ I \ {i}
    for j ∈ I do
        Compute α_j(I)
    end for
end while
return R.
```

construction, it is clear that

$$\alpha_i(J) \leq \alpha_i(I), \quad I \subset J.$$

Therefore, if H_J is rejected, there exists an $i \in J$ such that $p_i \leq \alpha_i(J)$. As $\alpha_i(J) \leq \alpha_i(I)$, we must have $p_i \leq \alpha_i(I)$ for all $i \in I \subset J$. Therefore, H_i will be rejected by the closure principle.

Since closure procedures always control the FWER, we can conclude that graphical procedures also control the FWER strongly.