Discussion of "Permutation-Based True Discovery Guarantee by Sum Tests"

by Anna Vesely, Livio Finos, Jelle J. Goeman

Pierre Neuvial

CNRS and Institut de Mathématiques de Toulouse (France)

2021-02-25

Features that I liked

- generic treatment of sum tests
- use of permutations to adapt to dependence
- generic formulation of the post hoc bound

$$q(S) = \max\{z \in \{0, \ldots, s+1\}, \phi(z) = 0\}$$

- shortcut in near-linear time in the number of hypotheses (for a given S)
- beyond Simes-based post hoc inference
- a sum test after truncation is still a sum test!
- implementation (R package sumSome on github)

Comment: formulation of the post hoc bound

Closed testing

- $q(S) = \max_{Z \notin \mathcal{R}} |Z \cap S|$
- $\mathcal{Z}_z = \{Z \subseteq S : |Z| \geq z\}$
- $\phi(z) = 1\{\mathcal{Z}_z \subseteq \mathcal{R}\}$

Local tests

- $q(S) = \max_{Z \notin \mathcal{R}_{loc}} |Z \cap S|$
- $\mathcal{V}_z = \{V \subseteq M : |V \cap S| \geq z\}$
- $\phi(z) = 1\{\mathcal{V}_z \subseteq \mathcal{R}_{\text{loc}}\}$

Then

$$q(S) = \max\{z \in \{0, \ldots, s+1\}, \phi(z) = 0\} \quad (*)$$

In practice $\phi(z)$ is calculated based on the local test formula.

- Genovese & Wasserman, *Ann. Stat.*, 2006
- Hemerik, Goeman & Solari, *Biometrika* (2019)
- Blanchard, N. & Roquain, *Ann. Stat.* (2020)
- Katsevich & Ramdas, *Ann. Stat.* (2020)

Questions on the post hoc bound $q(S)$

> Q1. Do we really need closed testing here?

Very short proof for (*) using only local tests: call q' the max on the right hand side. By definition of ϕ:

- $\phi(q') = 0 \quad \Rightarrow \quad \exists V : |V \cap S| \geq q', V \notin \mathcal{R}_{\text{loc}} \quad \Rightarrow \quad q(S) \geq q'$
- $\phi(q'+1) = 1 \quad \Rightarrow \quad \forall V : |V \cap S| \geq q'+1, V \in \mathcal{R}_{\text{loc}} \quad \Rightarrow \quad q(S) < q'+1$

> Q2. The formula

$$q(S) = \max\{z \in \{0, \dots, s+1\}, \phi(z) = 0\}$$

is not specific to sum test. Can it be/has is been used to derive shortcuts for other local tests?

Question on the shortcuts

Worst case complexity of the single-step shortcut for a given S:

$$O(Bm\log^2(m))$$

Iterated shortcut

Result: $d^{(n)}(S) = d(S)$ after at most $2^m - 2$ steps (!).

- n governs both the tightness of the bound and the time complexity.

> Q3. Can we characterize situations in which the single-step shortcut is exact, ie $q_0(S) = q(S)$?
>
> - further assumptions on the test statistics?

> Q4. Empirical study of which parameters influence n?
>
> - signal strength, proportion of non null items, number of tests…

Related works on permutation-based true discovery guarantees

> Q5. How does the prosed bounds compare to other permutation-based post hoc bounds?

- Andreella, Hemerik, Weeda, Finos, & Goeman, arxiv:2012.00368 (2020)
- Blanchard, N. & Roquain *Ann. Stat.* (2020)
- Blanchard, N. & Roquain, Book chapter for *Handbook of multiple testing* (to appear)

Recap of questions

> Q1. Do we need closed testing here?

> Q2. Could the expression
>
> $$q(S) = \max\{z \in \{0, \ldots, s+1\}, \phi(z) = 0\}$$
>
> be useful beyond sum tests?

> Q3. Further assumptions leading to exact single-step shortcut?

> Q4. Empirical study of which parameters influence n?

> Q5. Comparison to other permutation-based post hoc bounds