Online Control of the False Discovery Rate under “Decision Deadlines”

Aaron Fisher

Foundation Medicine Inc.; 150 Second St, Cambridge, MA 02141. afishe27@alumni.jh.edu; aaronjfisher.github.io.

Background & goal

Online testing methods let us make real time decisions about sequences of hypotheses.

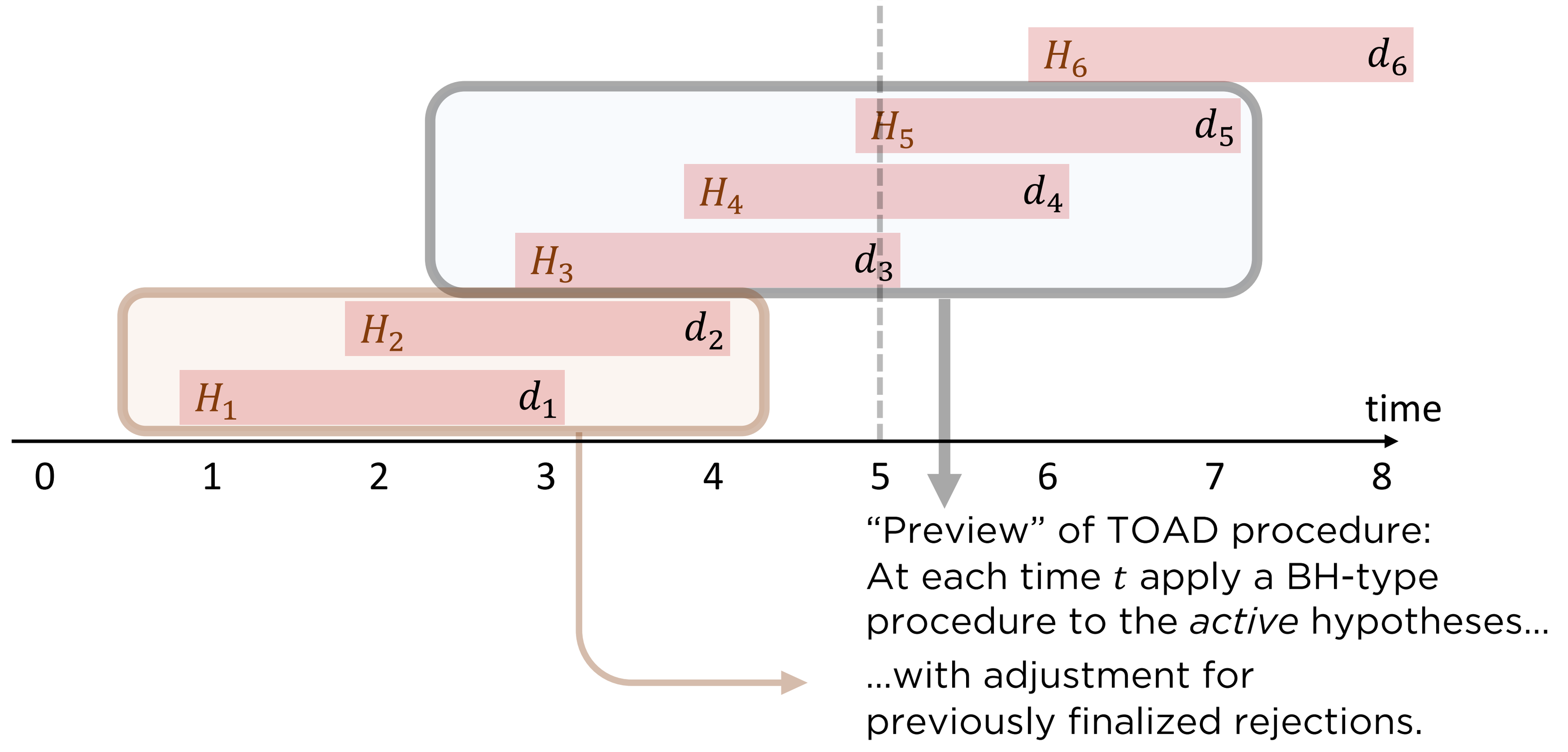
- E.g., fraud detection; sequential treatment assignment.

Goal: generate preliminary decision about each hypothesis in real time, but allow decision updates up until some present deadline.

- E.g., publishing results from sequences of experiments, with the option to retract or supplement in the future.

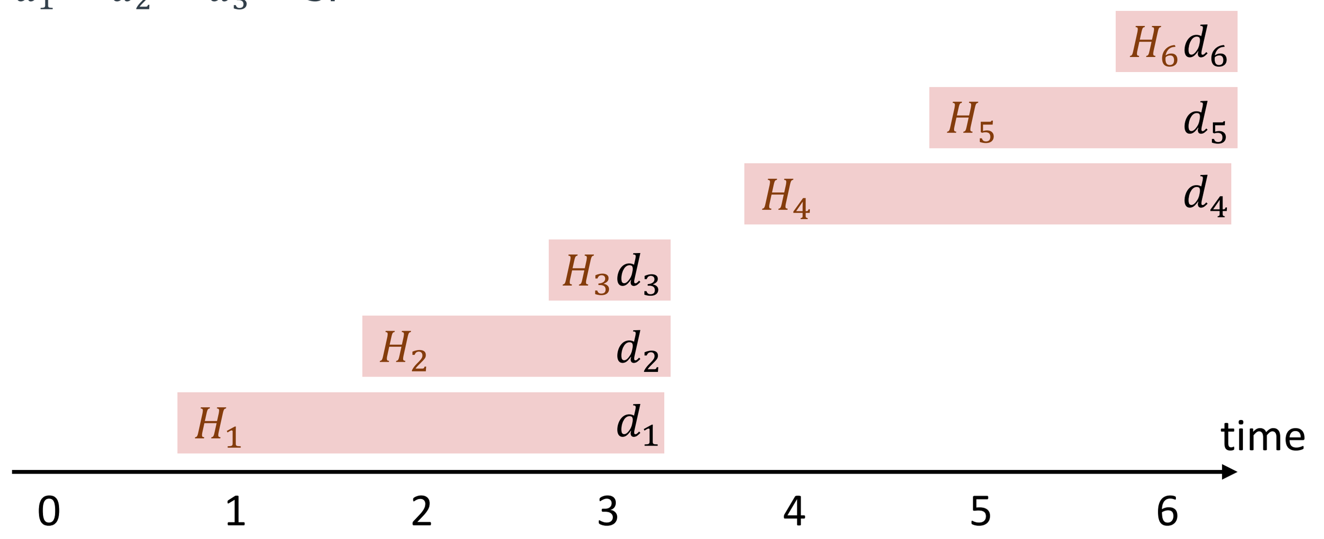
Ex1: moving window

- Final rejection decisions must be made within w stages (e.g., $w = 2$).
- That is, the decision deadline d_i for hypothesis H_i is $d_i := i + w$.



Ex2: Batch testing

- Hypotheses $H_1, H_2, \dots$ are grouped in batches of size 3.
- Final decisions are due at the end of each batch
  - e.g., $d_1 = d_2 = d_3 = 3$.



While some methods exist for the special case of batch testing (Zrnic et al., 2020), these methods are either less powerful or require more assumptions than our proposed method.

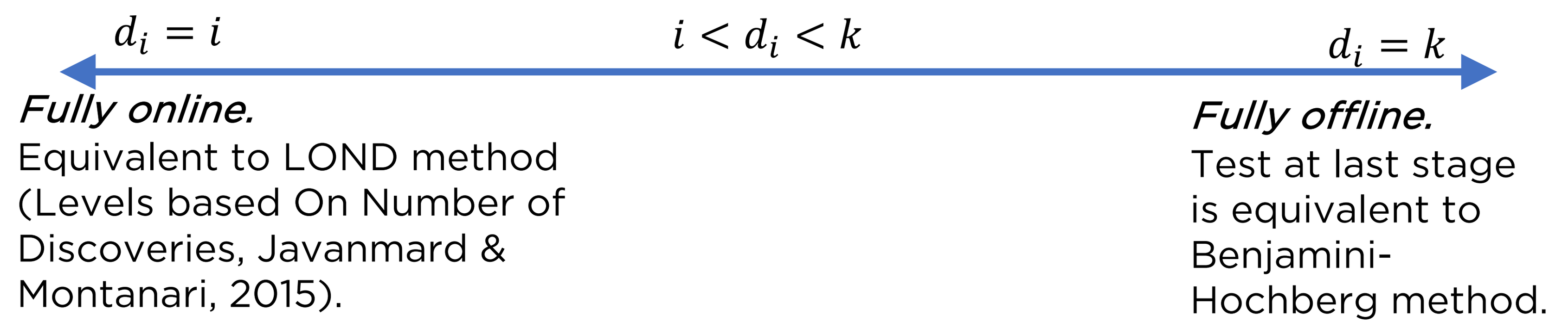
Algorithm: Thresholds Based on Active Discoveries (TOAD)

Below we show the simplest implementation of our algorithm, for a sequence of k hypotheses.

At each stage t:

1. *Let F_t be the current number of finalized rejections.*
2. *Let n_t be the hypotheses being **actively tested***
3. *Reorder these n_t hypotheses as $P_{(1,t)}, \dots P_{(n_t,t)}$.*
4. *Reject all p-values $\leq c_t$, where $c_t = max\{j \leq n_t : P_{(j,t)} \leq \frac{\alpha}{k}(j + F_t)\}$.*

Can be thought of as blending online & offline approaches:



Summary of analytical results

<u>FDR Control:</u>

- Under a positive dependence condition,

$$\mathrm{FDR}(t) = E\left[\frac{\#\text{ false discoveries by } t}{\max(1, \#\text{total discoveries by } t)}\right] \leq \alpha$$

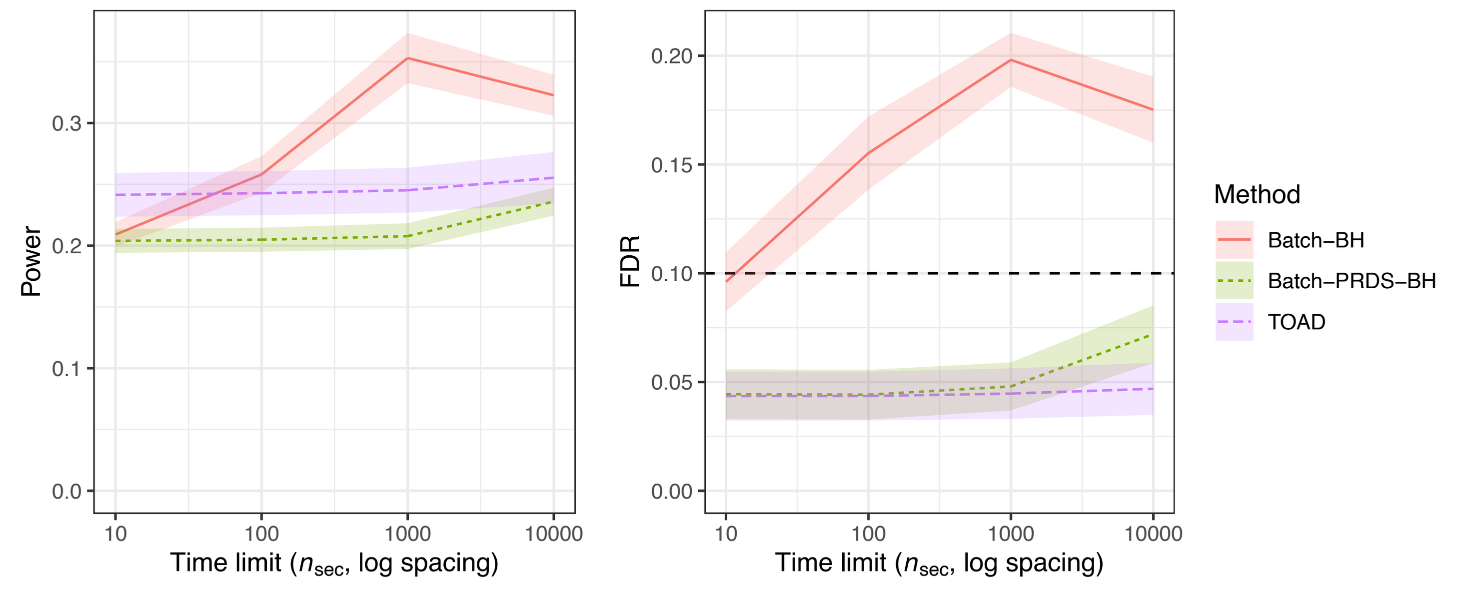
  for all t.
- Under arbitrary dependence, we can maintain FDR(t) control by incorporating shape functions
  - Analogous to the correction introduced by Benjamini and Yekutieli (2001).
- Under independence, $\mathrm{FDR}(T) \leq \alpha$ at *adaptively chosen stopping times T*.

<u>General behavior:</u>

- TOAD never reverses a previous rejection.
- TOAD is guaranteed to be at least as powerful as BATCH-BH-PRDS (Zrnic et al., 2020).
  - In our simulations, it is more powerful.

Applied example: credit card transactions

- Kaggle dataset of 284,807 credit card transactions over 48 hours (Pozzolo et al., 2015).
- Generate p-values for each transaction to flag fraud
- Apply online methods to the sequence
- Test decision deadlines ranging from 10 sec to ~3 hours



Of methods that control FDR at desired level, TOAD achieves best power

Overall summary

Our method, Thresholds Based on Active Discoveries (TOAD):

- allows users to choose decision deadlines for each test;
- controls FDR at each stage of testing;
- is more powerful than comparable methods that make the same assumptions; and
- under certain conditions, allows for adaptive stopping times.

Long term goal: Develop methods to study sequences of experiments, and the library of results they produce.

References

- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B Stat. Methodol., 57(1):289–300.
- Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Ann. Stat., 29(4):1165–1188.
- Javanmard, A. and Montanari, A. (2015). On online control of false discovery rate. arXiv [stat.ME].
- Pozzolo, A. D., Caelen, O., Johnson, R. A., and Bon- tempi, G. (2015). Calibrating probability with undersampling for unbalanced classification. In 2015 IEEE Symposium Series on Computational Intelligence, pages 159–166. ieeexplore.ieee.org.
- Zrnic, T., Jiang, D., Ramdas, A., and Jordan, M. (2020). The power of batching in multiple hypothesis testing. In Chiappa, S. and Calandra, R., editors, Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics, volume 108 of Proceedings of Machine Learning Research, pages 3806–3815. PMLR.