

## CONTROLLING THE FALSE DISCOVERY RATE THROUGH MULTIPLE COMPETITION

KRISTEN EMERY

(Received 10 October 2020; first published online 11 December 2020)

2020 *Mathematics subject classification*: primary 62F03; secondary 62G10, 62J05, 62P10.

*Keywords and phrases*: false discovery rate control, knockoffs, multiple hypotheses testing, target–decoy competition.

Multiple hypotheses testing is an important area of statistical research dedicated to the simultaneous analysis of multiple hypotheses. Error control in these testing scenarios is crucial to obtaining statistical rigour when deciding which of these hypotheses should be rejected. The most popular measure of this error is the false discovery rate (FDR). While there are many established procedures for obtaining guaranteed FDR control, most require the use of  $p$ -values which in some testing settings cannot be calculated. Competition-based FDR control provides a path to solving multiple hypotheses testing problems in these settings for which the canonical procedures fail.

As their name suggests, competition-based procedures use direct head-to-head competition between the originally observed score and a randomly generated null score for each hypothesis in order to control the FDR amongst the resulting list of discoveries. In this thesis we extend this competition framework and develop multiple testing procedures that utilise competition between the original scores and a multiple, rather than a single, set of null scores.

We construct these methods in the frameworks of peptide identification through target–decoy competition and the classical linear regression problem with Barber and Candès' recent knockoff procedure [1]. In both these cases we show through simulations and real data experiments that utilising multiple competition properly can lead to significant power gains without losing FDR control.

Some of the research in this thesis has been published in [2–4].

---

Thesis submitted to the University of Sydney in April 2020; degree approved on 6 August 2020; supervisor Uri Keich.

© 2020 Australian Mathematical Publishing Association Inc.

## References

- [1] R. F. Barber and E. J. Candès, ‘Controlling the false discovery rate via knockoffs’, *Ann. Statist.* **43**(5) (2015), 2055–2085.
- [2] K. Emery, S. Hasam, W. S. Noble and U. Keich, ‘Multiple competition based FDR control’, Preprint, 2019, [arXiv:1907.01458](https://arxiv.org/abs/1907.01458).
- [3] K. Emery, S. Hasam, W. S. Noble and U. Keich, ‘Multiple competition-based FDR control and its application to peptide detection’, in: *Research in Computational Molecular Biology*, Lecture Notes in Computer Science, 12074 (ed. R. Schwartz) (Springer, Cham, 2020).
- [4] K. Emery and U. Keich, ‘Controlling the FDR in variable selection via multiple knockoffs’, Preprint, 2019, [arXiv:1911.09442v2](https://arxiv.org/abs/1911.09442v2).

KRISTEN EMERY, School of Mathematics and Statistics,  
The University of Sydney, Camperdown,  
New South Wales 2006, Australia  
e-mail: [emery.kristen@hotmail.com](mailto:emery.kristen@hotmail.com)