

*XIV Congreso Galego de Estatística e Investigación de Operacións*  
Vigo, 24, 25 e 26 de outubro de 2019

## IMPROVED $Q$ -VALUES FOR DISCRETE AND HOMOGENEOUS TESTS

Marta Cousido-Rocha<sup>1</sup>, Jacobo de Uña-Álvarez<sup>1</sup>, Sebastian Döhler<sup>2</sup>

<sup>1</sup>Department of Statistics and Operations Research, SiDOR Research Group, and Centro de Investigaciones Biomédicas (CINBIO), Spain

<sup>2</sup>CCSOR and Faculty of Mathematics and Science, University of Applied Sciences Darmstadt, Germany

### ABSTRACT

Large scale discrete uniform homogeneous  $p$ -values arise in many applications, for example, in genome wide association studies. In this work several multiple testing procedures for such  $p$ -values are compared through simulations. Specifically, we consider the  $q$ -value approach based on five different estimators for the proportion of true null hypotheses  $\pi_0$ . Two of them are classical estimators of  $\pi_0$  for continuous  $p$ -values whereas the remaining three estimators are designed for discrete uniform homogeneous  $p$ -values. The  $q$ -value methods based on two of the latter estimators are novel in the multiple testing literature. The main simulated scenario is that of the two sample problem with low sample size and a large number of variables. Additional simulations based on the one sample problem scenario are also carried out. The main conclusion of this piece of work is that the discrete corrections may greatly improve the power of the  $q$ -values. An illustration with real data is included.

**Keywords:** Multiple testing procedures; Discrete  $p$ -values; High-dimensional data; Homogeneous  $p$ -values.

### ACKNOWLEDGMENTS

This work has received financial support of the Call 2015 Grants for PhD contracts for training of doctors of MINECO, cofinanced by FEDER,UE (Ref. BES-2015-074958). The authors acknowledge support from MTM2014-55966-P (MINECO) and MTM2017-89422-P (MINECO/AEI/FEDER,UE) projects. We also acknowledge the financial support provided by “Xunta de Galicia” (ED431C 2016/040).

### REFERENCES

- Chen, X., R. W. Doerge, and J. F. Heyse (2014). Methodology Multiple testing with discrete data: proportion of true null hypotheses and two adaptive FDR procedures. arXiv:1410.4274v2 .
- Cousido-Rocha, M., J. de Uña-Álvarez, and S. Döhler (2019a). DiscreteQvalue: Improved  $q$ -values for discrete uniform and homogeneous tests. R package version 1.0.
- Cousido-Rocha, M., J. de Uña-Álvarez, and J. Hart (2019b). A two-sample test for the equality of univariate marginal distributions for high-dimensional data. *Journal of Multivariate Analysis*, <https://doi.org/10.1016/j.jmva.2019.104537>.
- Habiger, J. D. (2015). Multiple test functions and adjusted  $p$ -values for test statistics with discrete distributions. *Journal of Statistical Planning and Inference* 167, 1-13.
- Kulinskaya, E. and A. Lewin (2009). On fuzzy familywise error rate and false discovery rate procedures for discrete distributions. *Biometrika* 96, 201-211.
- Liang, K. (2016). False discovery rate estimation for large scale homogeneous discrete  $p$ -values. *Biometrics* 72, 639-648.
- Storey, J. (2003). The positive false discovery rate: a bayesian interpretation and the  $q$ -value. *The Annals of Statistics* 31, 2013-2035.
- Storey, J. and R. Tibshirani (2003). Statistical significance for genomewide studies. *Proceedings of National Academy of Science* 100, 9440-9445.