False discovery rate control with multivariate *p*-values

Zhiyi Chi, *University of Connecticut*

Abstract

Multivariate statistics are often available as well as necessary in hypothesis tests. We study how to use such statistics to control not only false discovery rate (FDR) but also positive FDR (pFDR) with good power. We show that FDR can be controlled through nested regions of multivariate *p*-values of test statistics. If the distributions of the test statistics are known, then the regions can be constructed explicitly to achieve FDR control with maximum power among procedures satisfying certain conditions. On the other hand, our focus is where the distributions are only partially known. Under certain conditions, a type of nested regions are proposed and shown to attain (p)FDR control with asymptotically maximum power as the pFDR control level approaches its attainable limit. The procedure based on the nested regions is compared with those based on other nested regions that are easier to construct as well as those based on more straightforward combinations of the test statistics.

AMS 2000 subject classifications: Primary 62G10, 62H15; secondary 62G20.

Keywords: Multiple hypothesis testing, pFDR.

Full Text: PDF

References

[1]   Amit, Y. and Geman, D. (1997). Shape quantization and recognition with randomized trees. Neural Computation 9, 7, 1545–1588.

[2]   Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Stat. Methodol. 57, 1, 289–300. MR1325392

[3]   Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Ann. Statist. 29, 4, 1165–1188. MR1869245

[4]   Benjamini, Y. and Yekutieli, D. (2005). False discovery rate-adjusted multiple confidence intervals for selected parameters. J. Amer. Statist. Assoc. 100, 469, 71–93. With comments and a rejoinder by the authors. MR2156820

[5]   Blanchard, G. and Geman, D. (2005). Hierarchical testing designs for pattern recognition. Ann. Statist. 33, 3, 1155–1202. MR2195632

[6]   Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). Classification and regression trees. Wadsworth Statistics/Probability Series. Wadsworth Advanced Books and Software, Belmont, CA. MR0726392

[7]   Chi, Z. (2007). On the performance of FDR control: constraints and a partial solution. Ann. Statist. 35, 4, 1409–1431. MR2351091

[8]   Chi, Z. and Tan, Z. (2008). Positive false discovery proportions: Intrinsic bounds and

adaptive control. Statistica Sinica 18, 3, to appear.

[9]   Efron, B. (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. J. Amer. Statist. Assoc. 99, 465 (Mar.), 96–104. MR2054289

[10]   Efron, B., Tibshirani, R., Storey, J. D., and Tusher, V. G. (2001). Empirical Bayes analysis of a microarray experiment. J. Amer. Statist. Assoc. 96, 456, 1151–1160. MR1946571

[11]   Finner, H. and Roters, M. (2001). On the false discovery rate and expected type I errors. Biometri. J. 43, 985–1005. MR1878272

[12]   Genovese, C. and Wasserman, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. J. R. Stat. Soc. Ser. B Stat. Methodol. 64, 3, 499–517. MR1924303

[13]   Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control. Ann. Statist. 32, 3, 1035–1061. MR2065197

[14]   Meinshausen, N. and Rice, J. (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. Ann. Statist. 34, 1, 373–393. MR2275246

[15]   Munkres, J. R. (1984). Elements of algebraic topology. Addison-Wesley Publishing Company, Menlo Park, CA. MR0755006

[16]   Perone Pacifico, M., Genovese, C., Verdinelli, I., and Wasserman, L. (2004). False discovery control for random fields. J. Amer. Statist. Assoc. 99, 468, 1002–1014. MR2109490

[17]   R Development Core Team. (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

[18]   Signorovitch, J. E. (2006). Multiple testing with an empirical alternative hypothesis. Tech. Rep. 60, Harvard University Biostatistics Working Paper Series, Boston.

[19]   Spall, J. C. (2003). Introduction to stochastic search and optimization. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ. Estimation, simulation, and control. MR1968388

[20]   Storey, J. D. (2002). A direct approach to false discovery rates. J. R. Stat. Soc. Ser. B Stat. Methodol. 64, 3, 479–498. MR1924302

[21]   Storey, J. D. (2007). The optimal discovery procedure: a new approach to simultaneous significance testing. J. R. Stat. Soc. Ser. B Stat. Methodol. 69, 1, 1–22. MR2323757

[22]   Storey, J. D., Taylor, J. E., and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. J. R. Stat. Soc. Ser. B Stat. Methodol. 66, 1, 187–205. MR2035766

[23]   van der Laan, M. J., Dudoit, S., and Pollard, K. S. (2004). Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives. Stat. Appl. Genet. Mol. Biol. 3, Art. 15, 27 pp. (electronic). MR2101464

[24]   Vapnik, V. N. (2000). The nature of statistical learning theory, Second ed. Statistics for Engineering and Information Science. Springer-Verlag, New York. MR1719582