Copulas in multiple testing: Controlling the false discovery rate

Roy Cerqueti[1] **and Claudio Lupi**[2]
[1] University of Macerata, Italy
[2] University of Molise, Italy

This paper contributes to the development of a unifying perspective between copulas and *false discovery rate* (FDR) *control* — the control of the expected proportion of errors among the rejected hypotheses — in multiple testing with dependent test statistics. Indeed, despite the fact that test statistics dependence plays a fundamental role in multiple testing, the possibility of using copulas to characterize this dependence in a multiple testing framework remains largely unexplored: only Cerqueti et al. (2011, 2012), Dickhaus and Gierl (2012) and Bodnar and Dickhaus (2013) studied this topic in some detail.

The statistical relevance of FDR control procedures is grounded on the fact that the risk of falsely rejecting true null hypotheses increases dramatically when many hypotheses are tested simultaneously. Procedures which control the FDR ensure that the fraction of falsely rejected hypotheses over total rejections remains on average below a pre-specified level $q \in (0, 1)$. Benjamini and Hochberg (1995) proposed the first such procedure for independent test statistics, which was labelled BH after the names of its authors. Later, a number of contributions appeared in the literature dealing with FDR control under dependence assumptions. In particular, Benjamini and Yekutieli (2001) proved that the original BH procedure can be used to control the FDR not only in the presence of independent test statistics, but also in the case of positively dependent ones. Lately, Yekutieli (2008) proposed the general but less conservative *separate subsets BH* (ssBH) procedure.

We develop formal arguments to characterize the relationship between FDR control and the dependence properties of the underlying individual test statistics, when either the BH or the ssBH procedure is used. To achieve our aim, we model the stochastic dependence among the test statistics through the introduction of suitable families of copulas.

It is worth noting that there are several potential benefits from employing copulas in a multiple testing framework to describe the dependence structure of the individual test statistics:

- the introduction of copulas leads to dependence concepts more general than Pearson's correlation or those based on linearity;
- Sklar's Theorem (Sklar, 1959) allows one to obtain the joint distribution of a multivariate test through copulas by the knowledge of the marginals;
- the adoption of copulas improves parameters estimation procedures;

- dependence structures described by copulas are quite easy to be identified.

In order to let the analysis be as general as possible, we consider a non-exchangeable set of test statistics, allowing for exchangeability only within peculiar subsets of the statistics. More specifically, we start from the bivariate symmetric case; then we move to the case where exchangeability is allowed only within subsets of the test statistics corresponding to the true null hypotheses; finally, we generalize the discussion with reference to a family of multivariate asymmetric copulas. Hence, the framework adopted in the paper allows us to deal with FDR control in multiple testing in the relevant case of *heterogeneity*, to be intended as highly diversified dependence structures and coexistence of positive and negative pairwise correlations between the involved individual test statistics.

Keywords: False discovery rate, copulas, multiple testing, dependent test statistics.

References:

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B 57(1)*, 289-300.

Benjamini, Y., Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics 29(4)* 1165-1188.

Bodnar, T., Dickhaus, T. (2013). False discovery rate control under Archimedean copula. *arXiv preprint arXiv:1305.3897*, Humboldt-University Berlin.

Cerqueti, R., Costantini, M., Lupi, C. (2011). FDR control in the presence of an unknown correlation structure. *Economics & Statistics Discussion Paper 059/11*, University of Molise.

Cerqueti, R., Costantini, M., Lupi, C. (2012). A copula-based analysis of false discovery rate control under dependence assumptions. *Economics & Statistics Discussion Papers 065/12*, University of Molise.

Dickhaus, T., Gierl, J. (2012). Simultaneous test procedures in terms of p-value copulas. *SFB 649 Discussion Paper 2012-49*, Humboldt University, Berlin.

Sklar, A. (1959). Fonctions de répartition á n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris 8*, 229-231.

Yekutieli, D. (2008). False discovery rate control for non-positively regression dependent test statistics. *Journal of Statistical Planning and Inference 138(2)*, 405-415.