# Copulas, Uncertainty, and False Discovery Rate Control

Roy Cerqueti<sup>1 2</sup>

Claudio Lupi $^3$ 

University of Macerata

University of Molise

June 6, 2018

<sup>1</sup>University of Macerata, Department of Economics and Law. Via Crescimbeni, 20. I-62100 Macerata (Italy). E-mail: roy.cerqueti@unimc.it

<sup>2</sup>Corresponding author

<sup>3</sup>University of Molise, Department of Economics, Via F. De Sanctis. I-86100 Campobasso (Italy). E-mail: lupi@unimol.it

#### Abstract

The false discovery rate (FDR) is a powerful approach to multiple testing. However, dependence among test statistics is critical for FDR control. The way in which this dependence structure is described represents the most prominent source of uncertainty of this statistical theme. copulas play a relevant role among the techniques used to deal with uncertainty and dependence. This paper contributes to fill an existing gap in the scientific debate by exploring the connections between the literature on FDR and that on copulas. In particular, we aim at attracting the interest of the scientific community on this topic by identifying suitable classes of nonstandard copulas which ensure that FDR control can be attained for dependent test statistics.

*Key words*: Copulas; Uncertainty; Dependent test statistics; Multivariate total positivity of order 2; False discovery rate; Multiple testing.

#### 1 Introduction

In a multiple testing framework, *false discovery rate control* (FDR) amounts to controlling the expected proportion of errors among the rejected hypotheses. It was originally proposed by Benjamini and Hochberg (1995) and it has become central in the multiple testing literature (see, e.g., Yekutieli and Benjamini, 1999; Benjamini and Yekutieli, 2001; Storey, 2002; Sarkar, 2002, 2004; Farcomeni, 2007; Wu, 2008; Guo and Rao, 2008; Bodnar and Dickhaus, 2014, to cite a few). Recent applications of the FDR span a vast spectrum of fields (see, e.g., Pawluk-Koc et al., 2006; Bonissone et al., 2010; Fischer and Wermers, 2013; Glickman et al., 2014; Li et al., 2018). Despite its importance and diffusion in the theoretical as well in the applied literature, the FDR is far from being the only approach to multiple testing: other approaches have been advocated in the literature, including control of the familywise error rate (FWER: see, e.g., Simes, 1986; Sarkar and Chang, 1997; Lehmann and Romano, 2005), resampling-based procedures (Westfall and Young, 1993), and interesting recent contributions related to prior-free probabilistic inference (see, e.g., Dempster, 2008; Balch, 2012; Liu and Xie, 2014b,a).

This paper contributes to the development of a unifying perspective between copulas (see, e.g., Nelsen, 2006) and FDR control in multiple testing with dependent test statistics. In particular, we deal with one of the main source of uncertainty in multiple testing and describe the dependence among  $m \gg 1$  test statistics through copulas. Specifically, we identify wide families of copulas such that the FDR control can be achieved by using existing procedures.

Copulas have been widely used for modelling the stochastic dependence among random quantities, mainly in the context of empirical applications (see, e.g., Jondeau and Rockinger, 2006; Zimmer and Trivedi, 2006; Sriboonchitta et al., 2017). This explains the interest for this methodological tool under a theoretical point of view, with a growing interest in constructing new families of copulas and assessing their features (see, e.g., Demarta and McNeil, 2005; Durante et al., 2008; Mesiar et al., 2010; Durante et al., 2017; Wisadwongsa and Tasena, 2018). However, in the specific framework of multiple testing — where dependence plays a fundamental role — until recently the use of copulas has been only barely mentioned by Sarkar

(2008b) and remains largely unexplored or, at least, not explicitly dealt with. Dickhaus and Gierl (2013) and Stange et al. (2015) studied the use of copulas with reference to the FWER control; Schmidt et al. (2014, 2015) dealt with copulas for the control of the type I error rates in the context of adaptive designs; Cerqueti et al. (2011, 2012) and Bodnar and Dickhaus (2014) investigated the use of copulas for FDR control. However, Dickhaus and Gierl (2013) and Bodnar and Dickhaus (2014) adopt two perspectives which are radically different from the target of the present paper, which is the critical description of the dependence among test statistics in the light of FDR control. In fact, Bodnar and Dickhaus (2014) deal with the assessment of the upper and lower bounds for the FDR when the stochastic dependence is captured by an Archimedean copula, whereas Dickhaus and Gierl (2013) investigate simultaneous testing procedures that are able to control the FWER under some conditions and provide a representation of the common critical testing threshold on the basis of the quantiles of a special family of copulas. Finally, it is also important to give credit to He et al. (2017), where an estimation procedure of a high dimensional Gaussian graphical copula model is proposed. The high dimension of the problem leads to a multiple testing-FDR control issue, which is efficiently dealt with. However, the subject of that paper is again far from what we aim at dealing with here.

The statistical relevance of FDR control procedures is grounded on the fact that the probability of falsely rejecting true null hypotheses increases way over the nominal significance level  $\alpha$  when many hypotheses are simultaneously tested. On the other hand, generally a researcher would like to identify as many true rejections (discoveries) as possible, while incurring in a small proportion of false positives. Statistical procedures that control the FDR ensure that the fraction of falsely rejected hypotheses over total rejections — i.e., the false positive proportion — remains on average below a pre-specified level  $q \in (0, 1)$ . Benjamini and Hochberg (1995) proposed the first procedure of this kind for independent test statistics, which was labelled BH after the name of its authors. Later, 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. Building on Benjamini and Yekutieli (2001), a number of papers dealing with FDR control under dependence appeared: a necessarily non exhaustive list includes Kwong et al. (2002), Sarkar (2002), Cai and Sarkar (2006), Yekutieli (2008). In particular, the latter proposed a direct generalization of the BH method, called *separate subsets* BH (ssBH) procedure. The common idea contained in these papers is that the test statistics under the null have a form of dependence that can be described as PRDS (positive regression dependence on each one from a subset) or MTP2 (multivariate total positivity of order 2), where the latter implies the former.

In this paper we start from the same viewpoint, but we deal with the MTP2 property within the copula framework. The reason of this choice is motivated by the observation that the required MTP2 property has been proved only with reference to a limited number of multivariate distributions (see, e.g., Karlin and Rinott, 1980, 1981; Laradji, 2015, for some important contributions). In contrast, by using copulas we can generalize the type of distributions that can be considered by using general forms of dependence and/or nonstandard marginal distributions (e.g., Dickey-Fuller: Dickey and Fuller, 1979). Furthermore, it is important to notice that, even if all the test statistics under the null have the same marginal distribution (e.g., normal), the joint distribution is not generally the multivariate version of the marginals, therefore MTP2 multivariate distributions can be invoked only in very special circumstances. Copulas offer a flexible solution by separating the margins from the dependence structure, thus allowing the researcher to deal also with multivariate distributions having different and/or nonstandard marginals.

It is well known that permutation symmetry is a distinguishing feature of some classes of copulas that can be seen as a symmetry property. Such a property can be extended also to classes of quasi-copulas (see, e.g., Arias-García et al., 2017). In this respect, Archimedean copulas (for a survey we refer the reader to Schweizer and Sklar, 2005) and semilinear ones (see Durante, 2007; Durante et al., 2008) are important examples of permutation symmetric

copulas. However, in a multiple testing framework exchangeability may represent a severe restriction of the dependence structure among the individual test statistics. Hence, in order to make our analysis as general as possible, we discuss also the case of non-exchangeable sets of test statistics, allowing for exchangeability only within peculiar subsets of the statistics. 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. For some insights on asymmetric copulas and their applications, see e.g. Wei and Kim (2018); Wu (2014).

The rest of the paper is organized as follows. Section 2 gives preliminaries and notations. Section 3, the core of the paper, contains the formalization of the connection between FDR control and dependence of the test statistics through copulas. Section 4 collects a discussion of the theoretical results and draws some concluding remarks.

### 2 Preliminaries and notation

To let this paper be self-contained, we firstly recall two important definitions on the stochastic dependence which are useful in the FDR control framework.

**Definition 2.1.** Consider an increasing set<sup>1</sup> D. The m-variate random variable  $\boldsymbol{w}$  is assumed to satisfy the positive regression dependency on each one from a subset  $I_0 \subseteq \{1, \ldots, m\}$  (briefly, PRDS on  $I_0$ ) if, for each  $i \in I_0$ , the conditional probability  $\Pr(\boldsymbol{w} \in D | w_i = x)$  is nondecreasing in x.

**Definition 2.2.** Let f be the joint density function of the m-variate random variable  $\boldsymbol{w}$ . Then  $\boldsymbol{w}$  is said to satisfy the multivariate total positivity of order 2 (or to be MTP2) if and only if, for each  $\boldsymbol{x}$  and  $\boldsymbol{y}$  in  $\mathbb{R}^m$ , it results:

 $f(\boldsymbol{x}) \cdot f(\boldsymbol{y}) \leq f(\min\{\boldsymbol{x}, \boldsymbol{y}\}) \cdot f(\max\{\boldsymbol{x}, \boldsymbol{y}\})$ 

<sup>&</sup>lt;sup>1</sup>A set D is said to be increasing when, if  $x \in D$  and  $y \ge x$ , then  $y \in D$ .

where the min and max operators have to be intended componentwise. If m = 2, then MTP2 is briefly denoted as TP2 (total positivity of order 2).

A classical result in the theory of stochastic dependence states that if  $\boldsymbol{w}$  is MTP2, then it is also PRDS on each  $I_0$ . Moreover, if  $\{w_1, \ldots, w_m\}$  are MTP2, then the covariances  $\sigma_{i,j} \ge 0$ , for each  $i, j = 1, \ldots, m$ .

Let  $\mathbf{p} = (p_1, \ldots, p_m)'$  be the vector of the *m p*-values associated with the components of the collection of *m* test statistics  $\mathbf{t} = (t_1, \ldots, t_m)'$ . Consistently with Yekutieli (2008), we assume that the *p*-values in  $\mathbf{p}$  are obtained by applying a monotone transformation of the corresponding statistics in  $\mathbf{t}$ . This assumption ensures that if the test statistics are MTP2, so are the *p*-values (see, e.g., Sarkar, 2008a).

The BH procedure (Benjamini and Hochberg, 1995) is based on the sorted *p*-values  $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m-1)} \leq p_{(m)}$ . Let  $\mathbf{H}_{(0i)}$  be the null hypothesis corresponding to  $p_{(i)}$  and let k be the largest i for which  $p_{(i)} \leq \frac{i}{m}q$ . If k > 0, reject all  $\mathbf{H}_{(0i)}$   $i = 1, 2, \ldots, k$ . Benjamini and Hochberg (1995) showed that for independent test statistics and any configuration of false null hypotheses, the procedure controls the FDR at level q. Later, Benjamini and Yekutieli (2001) proved that it is not necessary that the test statistics are independent: rather, the PRDS property on the subset of the test statistics corresponding to the true null hypothesis ensures that the BH procedure controls the FDR at a level not greater than q. Furthermore, given that MTP2 implies PRDS on  $I_0 \forall I_0$ , the dependence described by the MTP2 property can be used instead of the PRDS on  $I_0$ .

Let us now divide p into S < m sub-vectors  $p^s$ , with s = 1, ..., S. The vector of p-values corresponding to the true null hypotheses will be denoted as  $p_0$ , and the p-values in  $p^s$  corresponding to the true null hypotheses will be denote as  $p_0^s$ . Let us assume that the number of components of  $p^s$ ,  $p_0$  and  $p_0^s$  are  $m^s$ ,  $m_0$  and  $m_0^s$ , respectively. With a very intuitive notation, the test statistics corresponding to  $p^s$ ,  $p_0$  and  $p_0^s$  constitute vectors, that will be indicated with  $t^s$ ,  $t_0$  and  $t_0^s$ , respectively.

The ssBH procedure (Yekutieli, 2008) runs into two steps as follows:

- 1. For s = 1, ..., S, apply the BH procedure at level  $qm^s/m$  to test  $\mathbf{p}^s$ , and denote as  $\mathbf{r}_{BH}^s$  the sets of *p*-values (components of  $\mathbf{p}^s$ ) corresponding to the rejected hypotheses.
- 2. Reject the null hypothesis corresponding to  $r_{ssBH} = \bigcup_{s=1}^{S} r_{BH}^{s}$ .

Yekutieli (2008) proved that this procedure controls the FDR at level  $q\frac{m_0}{m}$  when, for each  $s = 1, \ldots, S$ , the vector  $p^s$  is PRDS on  $p_0^s$ . This condition is implied by assuming that the  $p^s$ 's are MTP2. The fact that the MTP2 does not need to hold globally ensures that FDR control can be attained also in the presence of negative dependence, provided that the subsets are appropriately selected. Furthermore, the quoted paper also states that ssBH is less powerful than BH when S > 1 by construction — it coincides with BH when S = 1 — and this property is independent from the dependence structure among the statistics, under the obvious constraint that the FDR is controlled.

We deal here with MTP2 through the introduction of a very general dependence concept, the so-called *multivariate copula* (or, simply, *copula*). We report the formal definition of this instrument, and refer the interested reader to Nelsen (2006) for further details:

**Definition 2.3.** The function  $C : [0,1]^m \to [0,1]$  is a copula if and only if:  $(C2.3.i) \ C(u_1,\ldots,u_m) = 0$  if  $u_1 \times \ldots \times u_m = 0$ ;  $(C2.3.ii) \ C(u_1,\ldots,u_m) = u_{\bar{k}}$  if  $u_k = 1$ , for each  $k \neq \bar{k}$ ; (C2.3.iii) Given the m-dimensional rectangle  $[a_1,b_1] \times \ldots \times [a_m,b_m] \subseteq [0,1]^m$ , then

$$\sum_{i_1=1}^2 \dots \sum_{i_m=1}^2 (-1)^{i_1+\dots+i_m} C(u_{1,i_1},\dots,u_{m,i_m}) \ge 0,$$

where  $u_{j,1} = a_j$  and  $u_{j,2} = b_j$ .

The relevance of the concept of copula in describing stochastic dependence can be appreciated from Sklar's Theorem (Sklar, 1959). We adapt its enunciation to our specific notation set:

**Theorem 2.4.** (Sklar, 1959) Let  $F_{i_1,\ldots,i_{m^s}}$  be the joint distribution function of the  $m^s$ -ple  $(t_{i_1},\ldots,t_{i_{m^s}})$ , with  $i_1,\ldots,i_{m^s}=1,\ldots,m$ . Define the margins as  $F_{i_1},\ldots,F_{i_{m^s}}$ . Then there

exists a  $m^s$ -variate copula  $C_{i_1,\ldots,i_{m^s}}$  such that, for each  $x_1,\ldots,x_{m^s} \in \mathbb{R}$ ,

$$F_{i_1,\dots,i_{m^s}}(x_1,\dots,x_{m^s}) = C_{i_1,\dots,i_{m^s}}(F_{i_1}(x_1),\dots,F_{i_{m^s}}(x_{m^s})).$$
(1)

If the margins  $F_{i_1}, \ldots, F_{i_{m^s}}$  are continuous, then the copula  $C_{i_1,\ldots,i_{m^s}}$  is unique. Conversely, if  $C_{i_1,\ldots,i_{m^s}}$  is a  $m^s$ -variate copula and  $F_{i_1},\ldots,F_{i_{m^s}}$  are distribution functions, then the function  $F_{i_1,\ldots,i_{m^s}}$  defined in (1) is a  $m^s$ -dimensional distribution function with margins  $F_{i_1},\ldots,F_{i_{m^s}}$ .

Theorem 2.4 points out that, given a set of random variables, the relationship between joint and marginal distributions is stated through copulas.

The concept of permutation symmetric copulas plays a role also in this study. We report the related definition below for the convenience of the reader, and refer to Ghiselli Ricci (2013) for details:

**Definition 2.5.** The copula  $C : [0,1]^m \to [0,1]$  is permutation symmetric if, for each permutation  $\rho$  of  $\{1, \ldots, m\}$ , one has:

$$C(u_1,\ldots,u_m)=C(u_{\varrho(1)},\ldots,u_{\varrho(m)}).$$

Finally, for the sake of simplicity, in the rest of the paper we adopt set theory notation to describe the vectors  $p^s$ :

Notation 2.6. The p-value  $p_i$  is contained in  $\mathbf{p}^s$  — i.e.  $p_i \in \mathbf{p}^s$  — if and only if  $p_i$  is a component of the vector  $\mathbf{p}^s$ . Moreover,  $\mathbf{p}^s = \emptyset$  if and only if  $\nexists i \in \{1, \ldots, m\}$  such that  $p_i \in \mathbf{p}^s$ . The cardinality of  $\mathbf{p}^s$  is the number of its components.

#### 3 Main results

Although we focus on the ssBH procedure, it is worth noticing that all the results contained in this Section can be easily written in terms of the original BH procedure which — as already stated above — can be mathematically viewed as a particular case of the ssBH by assuming S = 1.

Our first remark concerns a particular family of permutation symmetric copulas which characterize the MTP2 property. Specifically, we firstly restrict to the important case of subsets of cardinality 2, and then extend to the general setting.

**Proposition 3.1.** Consider s = 1, ..., S such that  $m^s = 2$  and  $m_0^s > 0$ . Moreover suppose that, the dependence among the statistics in  $t^s$  is described by a copula  $C_s$  such that:

$$C_s(u_1, u_2) = u_1 u_2 + \theta_s \phi_s(u_1) \phi_s(u_2), \tag{2}$$

with  $\theta_s \in [-1, 1]$  and  $\phi_s : [0, 1] \rightarrow [0, 1]$  satisfying the following conditions:  $(C3.1.i) \ \phi_s(0) = \phi_s(1) = 0;$ 

(C3.1.ii)  $\phi_s$  is Lipschitzian in [0, 1], i.e.:  $|\phi_s(u_1) - \phi_s(u_2)| \le |u_1 - u_2|$ , for each  $u_1, u_2 \in [0, 1]$ ; (C3.1.iii)  $\phi_s$  is convex or concave in [0, 1].

Then the level q ssBH procedure controls the FDR at level  $qm_0/m$ .

*Proof.* Let  $t_i$  and  $t_j$  denote the individual statistics in  $t^s$ . Amblard and Girard (2002) shows that, if the dependence between  $t_i$  and  $t_j$  is described through the copula  $C_s$  in (2) and Condition (C3.1.iii) holds, then  $t_j$  is stochastically increasing in  $t_i$  and  $t_i$  is stochastically increasing in  $t_j$ , i.e. the following conditions hold:

$$\begin{cases}
P(t_j > x_j | t_i = x_i) \text{ is nondecreasing in } x_i, \forall x_j; \\
P(t_i > x_i | t_j = x_j) \text{ is nondecreasing in } x_j, \forall x_i.
\end{cases}$$
(3)

System (3) is equivalent to the TP2 property for the set  $t^s$  (see Nelsen, 2006). Being  $m_0^s > 0$ , this outcome leads to the thesis, in virtue of Proposition 2.2 in Yekutieli (2008).

The importance of the bivariate case is associated to the fact that the copula defined in (2) allows to derive an explicit expression for the correlation between the individual statistics in  $t^s$ .

Indeed, a straightforward computation provides that if the stochastic dependence between  $t_i$ and  $t_j$  is described through copula  $C_s$  in (2), then the Spearman's  $\rho$  between  $t_i$  and  $t_j$ , namely  $\rho_{i,j}$ , can be written as:

$$\rho_{i,j} = 12\theta_s \left(\int_0^1 \phi_s(\xi)d\xi\right)^2.$$
(4)

For a formal proof of the validity of (4), see Amblard and Girard (2002).

Notice also that copula (2) used in Proposition 3.1 may be viewed as a "perturbation" of the product copula: when  $\theta_s = 0$  the case collapses to independence. It is also worth noticing that copula (2) is a generalization of the permutation symmetric Farlie-Gumbel-Morgenstern (FGM) copula that holds when  $\phi_s(u) = u(1 - u)$ . However, a word of caution is in order here. The FGM copula, as well as its studied variants, are known for implying only modest dependence (see, e.g., Huang and Kotz, 1999): therefore, we cannot expect copula (2) to accurately represent very strong dependence across the test statistics. As far as the "pure" FGM copula is concerned, its dependence as measured by Kendall's  $\tau$  and Spearman's  $\rho$  is respectively  $2\theta_s/9$  and  $\theta_s/3$  with  $-1 \leq \theta_s \leq 1$ . However, an appropriate choice of a function  $\phi_s(u) \neq u(1-u)$  satisfying conditions in Proposition 3.1 allows to strengthen the typical weak dependence structure of the FGM copula.

Furthermore, a recent invariance result due to Durante et al. (2010) allows us to transform copula (2) by means of a suitable class of isomorphisms, again obtaining FDR control. In this respect, the family of copulas  $C_s$  in Proposition 3.1 may be interpreted as a *generator* for the dependence required to control the FDR.

We formalize this argument in the following Proposition:

**Proposition 3.2.** Consider s = 1, ..., S such that  $m^s = 2$  and  $m_0^s > 0$ . Moreover suppose that, for each s, the dependence among the statistics in  $\mathbf{t}^s$  is described by a copula  $C_s^{\xi}$  such that:

$$C_s^{\xi}(u_1, u_2) = \xi_s \left( C_s(\xi_s^{-1}(u_1), \xi_s^{-1}(u_2)) \right),$$
(5)

where  $C_s$  is defined as in (2) of Proposition 3.1 and  $\xi_s$ :  $[0,1] \rightarrow [0,1]$  is an increasing

bijection such that  $\xi_s \circ \exp: (-\infty, 0] \rightarrow [0, 1]$  is log-convex.

Then the level q ssBH procedure controls the FDR at level  $qm_0/m$ .

*Proof.* Since  $C_s$  in (2) is a copula, then the assumptions on function  $\xi_s$  assures that also  $C_s^{\xi}$  in (5) is a copula. Hence, Proposition 3.1 and Theorem 3.1 in Durante et al. (2010) assure that the dependence among the test statistics in  $t^s$  is of TP2-type, and Yekutieli (2008) gives the thesis.

Propositions 3.1 and 3.2 offer viable ways of selecting the couples in such a way that the conditions for validly using the ssBH procedure are satisfied. It is important to notice that the pairwise dependence introduced in the set up given by the family  $\{p^s\}_{s=1,...,S}$  and the definition of copulas (2) and (5) allow us to describe a system with both positively and negatively correlated test statistics. This aspect meets a natural requirement on the dependence structure of statistics in multiple testing.

There are two main limitations to the adoption of copulas (2) and (5). First, Propositions 3.1 and 3.2 refer to *couples*, i.e. subsets of cardinality 2; furthermore, as already stated, both copulas (2) and (5) are permutation symmetric, and this aspect may be a very strong requirement, in that symmetric copulas are able to represent only a small range of dependencies. However, it is noticing that permutation symmetry does not extend to the entire set of test statistics t, which in general may collect m non-exchangeable individual tests.

It is possible to have a more general view of the results above by using a  $m^s$ -variate approach, with  $m^s > 2$ , in a not necessarily permutation symmetric framework for the subsets of t. We first discuss the permutation symmetric  $m^s$ -variate case, with  $m^s > 2$ , and then provide a generalization to the nonsymmetric setting.

To deal with the  $m^s$ -variate symmetric framework, it is useful to recall a generalization of the monotonic property for functions:

**Definition 3.3.** A function

$$\psi^{-1}: [0, +\infty) \to [0, 1]$$
 (6)

is completely monotone in  $[0, +\infty)$  if and only if  $\psi^{-1} \in C^{\infty}(0, +\infty) \cap C^{0}[0, +\infty)$ , and  $(-1)^{n} (\psi^{-1})^{(n)} (x) \ge 0, \forall n = 0, 1, 2, ...; \forall x \in (0, +\infty).$ 

The copula of interest is the following:

**Definition 3.4.** Fix s = 1, ..., S and consider a function

$$\psi_s: [0,1] \to [0,+\infty) \tag{7}$$

such that  $\psi_s^{-1}$  exists and is completely monotone in  $[0, +\infty)$ ,  $\psi_s^{-1}(0) = 1$  and  $\lim_{x \to +\infty} \psi_s^{-1}(x) = 0$ . An  $m^s$ -variate Archimedean copula with generator  $\psi_s$  is a copula  $C_s^{(\psi)}$  such that

$$C_s^{\psi}(u_1, \dots, u_{m^s}) = \psi_s^{-1}\left(\sum_{i=1}^{m^s} \psi_s(u_i)\right).$$
 (8)

Schweizer and Sklar (2005) proved that, when  $m^s = 2$ , a copula as in (8) is Archimedean if and only if  $\psi_s^{-1}$  is convex. This fact does not extend to the case of  $m^s > 2$ . However, the condition of complete monotonicity for  $\psi_s^{-1}$  can be weakened, even in the case of  $m^s > 2$ . A sufficient condition for  $C_s^{\psi}$  to be a copula is to replace the complete monotonicity with the condition  $(-1)^n (\psi_s^{-1})^{(n)}(x) \ge 0$ , for each  $n = 0, 1, \ldots, m^s$  and  $x \in (0, +\infty)$ . For a detailed discussion on this topic, we refer to McNeil and Nešlehová (2009).

As we saw with reference to copulas (2) and (5), also the Archimedean copula  $C_s^{\psi}$  in (8) refers to the symmetric case of exchangeability for the subsets  $t^s$ . Such a copula describes the MTP2 property, hence being useful to control the FDR:

**Proposition 3.5.** Consider s = 1, ..., S such that  $m_0^s > 0$ . Assume that, the dependence among the statistics in  $\mathbf{t}^s$  is described by an Archimedean copula  $C_s^{\psi}$  as in (8), with:

$$u_k = F_k(x_k), \qquad x_k \in \mathbb{R}, \ \forall \, k = 1, \dots, m^s, \tag{9}$$

where  $(-1)^{m^s} (\psi_s^{-1})^{(m^s)}$  is log-convex.

Then the level q ssBH procedure controls the FDR at level  $qm_0/m$ .

Proof. Let  $t_1, \ldots, t_{m^s}$  denote the individual statistics in  $t^s$ . If the stochastic dependence in  $t^s$  is described as it was in the hypotheses, then Müller and Scarsini (2005) guarantees that the MTP2 property holds among the individual statistics in  $t^s$ . Hence, Proposition 2.2 in Yekutieli (2008) gives the result.

The need to obtain a truly multivariate result motivates the formulation of Proposition 3.5, even if permutation symmetry continues to be a feature of the class of multivariate Archimedean copulas involved. As with the bivariate case, it is worth noticing that the introduction of the family of sets  $\{p^s\}_{s=1,...,S}$  with the stochastic dependence structure formalized in Proposition 3.5 allows us to describe a non-exchangeable system with both positively and negatively correlated test statistics. In fact, the positive dependence condition modeled by permutation symmetric copula (8) is required only for some statistics in t (the ones collected in the  $t^s$ 's such that  $m_0^s > 0$ ). Furthermore, the Archimedean copulas used in Proposition 3.5 are more flexible than the particular case of FGM copulas, in that they can represent cases with both strong positive and negative dependence. In the case  $m^s = 2$ , Kendall's  $\tau_{i,j}$  for an Archimedean copula related to  $(t_i, t_j)$  takes the convenient form (see Genest and MacKay, 1986, Theorem 2).

$$\tau_{i,j} = 4 \int_0^1 \frac{\psi_s(y)}{\psi'_s(y)} dy + 1.$$
(10)

The generalization to the permutation asymmetric framework can be obtained at the cost of some mildly stronger assumptions, as recently explained by Cerqueti and Lupi (2016). We give full details by first introducing an asymmetric copula constituting a generalization of the Archimedean copula proposed in (8).

**Definition 3.6.** Fix s = 1, ..., S. Let us introduce  $J_s \in \mathbb{N}$  and a set of  $m^s \times J_s$  functions

$$h_{jks}: [0,1] \to [0,1], \qquad j = 1, \dots, J_s; \ k = 1, \dots, m^s$$
(11)

such that:

(C3.6.i)  $h_{jks}$  is differentiable in (0,1) and strictly increasing in [0,1], for each j, k;

 $(C3.6.ii) h_{jks}(0) = 0 \text{ and } h_{jk}(1) = 1, \text{ for each } j, k;$  $(C3.6.iii) \frac{1}{J_s} \sum_{j=1}^{J_s} h_{jks}(x) = x, \text{ for each } k = 1, \dots, m^s \text{ and } x \in [0, 1].$ Moreover, define

$$\psi_s : [0,1] \to [0,1]$$
 (12)

such that:

(C3.6.iv)  $\psi_s$  is  $m^s + 2$  times differentiable in (0,1); (C3.6.v)  $\psi_s^{(i)} > 0$ , for  $i = 1, ..., m^s$ ; (C3.6.vi)  $\psi_s(0) = 0$  and  $\psi_s(1) = 1$ .

We define a non-exchangeable (Archimedean) copula as  $C_{NE}^{\psi}: [0,1]^{m^s} \to [0,1]$  such that:

$$C_{NE}^{\psi}(u_1, \dots, u_{m^s}) = \psi_s^{-1} \left( \frac{1}{J_s} \sum_{j=1}^{J_s} \prod_{k=1}^{m^s} h_{jks}(\psi_s(u_k)) \right).$$
(13)

Copula (13) has been first introduced and explored by Liebscher (2008, 2011). It is worth noticing that, as far as the copula's definition is concerned, conditions (C3.6.iv) and (C3.6.v) could be weakened. However, the stronger conditions proposed here allow us to use Cerqueti and Lupi (2016, Theorem 3.3) and easily derive the following general result:

**Proposition 3.7.** Consider s = 1, ..., S such that  $m_0^s > 0$ . Suppose that the dependence among the statistics in  $\mathbf{t}^s$  is described by copula  $C_{NE}^{\psi}$  in (13). Furthermore, suppose that  $h_{jks}$ is twice differentiable in (0, 1), with

$$\prod_{k=k_1,k_2} \left[ h_{jks}''(\psi_s(u_k)) \left(\psi_s'(u_k)\right)^2 + h_{jks}'(\psi_s(u_k))\psi_s''(u_k) \right] \ge \left[ h_{jk_1s}''(\psi_s(u_{k_1})) \times \left(\psi_s'(u_{k_1})\right)^2 + h_{jk_1s}'(\psi_s(u_{k_1}))\psi_s''(u_{k_1}) \right] \times h_{jk_2s}'(\psi_s(u_{k_2}))\psi_s'(u_{k_2})$$
(14)

and

$$\prod_{k=k_1,k_2} h'_{jks}(\psi_s(u_k))\psi'_s(u_k) \ge \left[h''_{jk_1s}(\psi_s(u_{k_1}))(\psi'_s(u_{k_1}))^2 + h'_{jk_1s}(\psi_s(u_{k_1}))\psi''_s(u_{k_1})\right] \times h'_{jk_2s}(\psi_s(u_{k_2}))\psi'_s(u_{k_2}),$$
(15)

holds for each  $j = 1, \ldots, J$ ,  $k_1, k_2 = 1, \ldots, m$ ,  $k_1 \neq k_2$ , and  $s = 1, \ldots, S$ . Finally assume that, for each  $s = 1, \ldots, S$ , one has:

$$\left(\psi_{s}^{-1}\right)^{(m^{s}+2)}(x)\left(\psi_{s}^{-1}\right)^{(m^{s})}(x) - \left[\left(\psi_{s}^{-1}\right)^{(m^{s}+1)}(x)\right]^{2} \ge 0, \qquad \forall x \in (0,1).$$
(16)

Then the level q ssBH procedure controls the FDR at level  $qm_0/m$ .

*Proof.* We adapt to our case the arguments of Cerqueti and Lupi (2016) to prove that copula  $C_{NE}^{\psi}$  in (13) describes a dependence of type MTP2.

First we observe that a long but straightforward computation proves that the density f of  $C_{NE}^{\psi}$  is log-supermodular, i.e.

$$\log(f(\boldsymbol{u})) := \log\left(\frac{\partial^{m^s}}{\partial u_1 \dots \partial u_{m^s}} C_{NE}^{\psi_s}(\boldsymbol{u})\right)$$
(17)

is supermodular (for the details, refer to Cerqueti and Lupi, 2016).

Therefore, Müller and Scarsini (2005) guarantee that  $C_{NE}^{\psi}$  is MTP2.

Since  $m_0^s > 0$ , then Yekutieli (2008) implies that the level q ssBH procedure controls the FDR at level  $qm_0/m$ .

**Example 3.8.** It is now worth providing an example of the shape of a copula satisfying Proposition 3.7. To provide such an example, we deal with the bivariate case, i.e.:  $m^s = 2$ . Specifically, a simple computation gives that

$$C_{NE}^{\psi}(u_1, u_2) = \log\left[\frac{(e^{u_1} - 1)(e^{u_2} - 1)}{e - 1} + 1\right]$$
(18)

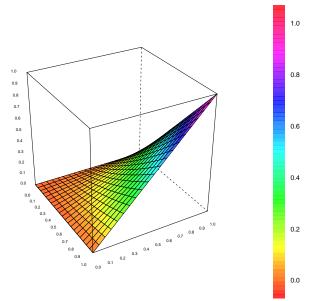


Figure 1: A graphical representation of copula (18)

belongs to the family defined in (13) and fulfils the hypotheses listed in Proposition 3.7. Indeed, fix s = 1, ..., S and  $J_s \in \mathbb{N}$ . It is sufficient to define

$$h_{j1}(x) = h_{j2}(x) = x, \qquad x \in [0, 1], \ j = 1, \dots, J_s,$$
(19)

and take

$$\psi_s(u_k) = \frac{e^{u_k} - 1}{e - 1},\tag{20}$$

for each k = 1, 2.

A graphical representation of copula (18) is offered in Figure 1. In order to give further details on this copula, we can also simulate from a bivariate distribution based on copula (18) with arbitrary marginal distributions. This can be achieved using conditional sampling, which is a very general algorithm for bivariate copulas. In particular, in Figure 2 we report the scatterplot of 10,000 samples from a joint distribution defined by copula (18) and t marginal distributions with 3 degrees of freedom.

Copulas (2) and (5), and the copula defined by (8)–(9) are permutation symmetric, in that they are invariant with respect to permutation of their univariate arguments. Such a

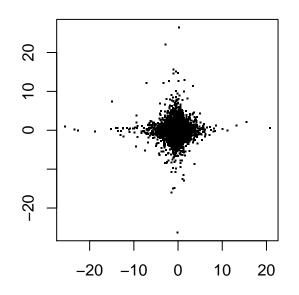


Figure 2: Scatterplot of 10,000 samples from a joint distribution defined by copula (18) and  $t_3$  marginal distributions.

symmetry is able to model a relatively small range of dependence structures on the involved subsets of t which depend only on a small number of parameters. This is their main limitation, implying that they are not particularly flexible in fitting multivariate data with a large number of parameters. However, it is worth stressing once again that exchangeability does not extend to the entire set t, which then may collect individual non-exchangeable test statistics.

For the sake of generality, we have provided some arguments on the extension to the nonexchangeable case. Proposition 3.7 extends the FDR applicability to situations where dependence can be well represented by asymmetric copulas. In this respect, Proposition 3.7 complements and extends Yekutieli (2008).

### 4 Discussion and concluding remarks

In multiple testing, a very relevant role is played by the analysis of the dependence structure of the involved test statistics. In doing so, the uncertainty of the model is efficiently dealt with. Copulas are particularly suitable for this purpose; nonetheless, the role of copulas in multiple testing has been scarcely investigated. This paper contributes to reduce this gap by critically illustrating the relationship between copulas and control of the false discovery rate.

Benjamini and Yekutieli (2001) and Yekutieli (2008) (among others) showed that, in the presence of dependent tests, a PRDS-type dependence among the test statistics under the null is sufficient for controlling the FDR. Since MTP2 is a stronger condition that implies PRDS, FDR is attained also if MTP2 holds.

A copula-based framework for the FDR control can bring three main advantages. First, the MTP2 property has been proved to hold with explicit reference to a rather narrow set of multivariate distributions, a prominent example being the multivariate normal distribution with certain types of non-negative covariances (Karlin and Rinott, 1980, 1981): on the other hand, copulas allow multiple testing under less restrictive assumptions on the multivariate distribution of the test statistics. Second, even if the marginal distributions of the test statistics under the null are known (e.g., normal), nevertheless the joint distribution of the test statistics may not obey one of the multivariate distributions known to be MTP2 (e.g., multivariate normal): through Sklar's theorem (Sklar, 1959), the copula approach deals with the margins and with the dependence structure separately, thus allowing for greater flexibility as far as the multivariate distribution is concerned, even permitting the use of test statistics having non-standard (marginal) distributions (e.g., unit root tests; Phillips, 1997). Third, Durante et al. (2010) proved that bivariate copulas maintain the TP2 property when they are suitably *distorted* (i.e., transformed through an opportunely defined isomorphism): hence copulas can be viewed as generators of couples of statistics which are MTP2, once a reference copula of MTP2 type is identified and suitably distorted.

We stress again that the copulas we discuss in the presented Propositions are strongly dependent on the index s = 1, ..., S. In this context, one could also check *a-priori* if a suitable collection of subsets of test statistics exists, whose dependence structure obeys to one of the copulas described in the present paper.

However, results remain valid, since the definition of ssBH procedure requires only the existence of a collection of subsets for FDR control — and so having the required dependence structure.

Moreover, this fact allows us to deal with FDR control in multiple testing in the relevant case of *heterogeneity* — to be intended as highly diversified dependence structures between the vectors  $t^s$  — of the involved individual test statistics. In particular, since we impose a copula-type dependence structure only to selected clusters of the *p*-values, and given that we consider also non-Archimedean and asymmetric Archimedean copulas, by using the copulabased theoretical framework we can represent not only situations in which the test statistics are exchangeable and positively dependent, but also cases with non-exchangeable and negatively correlated subsets of test statistics.

## References

- Amblard, C., Girard, S., 2002. Symmetry and dependence properties within a semiparametric family of bivariate copulas. Nonparametric Statistics 14 (6), 715–727.
- Arias-García, J., Mesiar, R., De Baets, B., 2017. The unwalked path between quasi-copulas and copulas: Stepping stones in higher dimensions. International Journal of Approximate Reasoning 80, 89–99.
- Balch, M. S., 2012. Mathematical foundations for a theory of confidence structures. International Journal of Approximate Reasoning 53 (7), 1003–1019.
- 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., 2014. False discovery rate control under Archimedean copula. Electronic Journal of Statistics 8 (2), 2207–2241.
- Bonissone, P., Cadenas, J. M., Garrido, M. C., Díaz-Valladares, R. A., 2010. A fuzzy random forest. International Journal of Approximate Reasoning 51 (7), 729 – 747.
- Cai, G., Sarkar, S. K., 2006. Modified Simes' critical values under positive dependence. Journal of Statistical Planning and Inference 136 (12), 4129–4146.

- 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. URL http://econpapers.repec.org/paper/molecsdps/esdp11059.htm
- 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. URL http://econpapers.repec.org/paper/molecsdps/esdp12065.htm

- Cerqueti, R., Lupi, C., 2016. Non exchangeable copulas and multivariate total positivity. Information Sciences 360, 163–169.
- Demarta, S., McNeil, A. J., 2005. The t copula and related copulas. International Statistical Review 73 (1), 111–129.
- Dempster, A., 2008. The Dempster–Shafer calculus for statisticians. International Journal of Approximate Reasoning 48 (2), 365–377.
- Dickey, D. A., Fuller, W. A., 1979. Distributions of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74 (366), 427–431.
- Dickhaus, T., Gierl, J., 2013. Simultaneous test procedures in terms of p-value copulae. In: Proceedings of the 2nd Annual International Conference on Computational Mathematics, Computational Geometry and Statistics (CMCGS 2013). Vol. 2. Global Science and Technology Forum (GSTF), pp. 75–80.
- Durante, F., 2007. A new family of symmetric bivariate copulas. Comptes Rendus Mathematique 344 (3), 195–198.
- Durante, F., Fernández-Sánchez, J., Quesada-Molina, J. J., Úbeda-Flores, M., 2017. Copulas with given values on the tails. International Journal of Approximate Reasoning 85, 59–67.
- Durante, F., Foschi, R., Sarkoci, P., 2010. Distorted copulas: Constructions and tail dependence. Communications in Statistics – Theory and Methods 39 (12), 2288–2301.
- Durante, F., Kolesárová, A., Mesiar, R., Sempi, C., 2008. Semilinear copulas. Fuzzy Sets and Systems 159 (1), 63–76.
- Farcomeni, A., 2007. Some results on the control of the false discovery rate under dependence. Scandinavian Journal of Statistics 34 (2), 275–297.
- Fischer, B. R., Wermers, R., 2013. Multiple fund performance evaluation: The false discovery rate approach. In: Fischer, B. R., Wermers, R. (Eds.), Performance Evaluation and Attribution of Security Portfolios. Academic Press, San Diego, Ch. 8, pp. 251–284.
- Genest, C., MacKay, J., 1986. The joy of copulas: Bivariate distributions with uniform marginals. The American Statistician 40 (4), 280–283.
- Ghiselli Ricci, R., 2013. Exchangeable copulas. Fuzzy Sets and Systems 220 (1), 88–98.

- Glickman, M. E., Rao, S. R., Schultz, M. R., 2014. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. Journal of Clinical Epidemiology 67 (8), 850–857.
- Guo, W., Rao, M. B., 2008. On control of the false discovery rate under no assumption of dependency. Journal of Statistical Planning and Inference 138 (10), 3176–3188.
- He, Y., Zhang, X., Wang, P., Zhang, L., 2017. High dimensional Gaussian copula graphical model with FDR control. Computational Statistics & Data Analysis 113, 457–474.
- Huang, J. S., Kotz, S., 1999. Modifications of the Farlie-Gumbel-Morgenstern distributions. A tough hill to climb. Metrika 49 (2), 135–145.
- Jondeau, E., Rockinger, M., 2006. The copula-garch model of conditional dependencies: An international stock market application. Journal of International Money and Finance 25 (5), 827–853.
- Karlin, S., Rinott, Y., 1980. Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. Journal of Multivariate Analysis 10 (4), 467–498.
- Karlin, S., Rinott, Y., 1981. Total positivity properties of absolute value multinormal variables with applications to confidence interval estimates and related probabilistic inequalities. Annals of Statistics 9 (5), 1035–1049.
- Kwong, K. S., Holland, B., Cheung, S. H., 2002. A modified Benjamini-Hochberg multiple comparisons procedure for controlling the false discovery rate. Journal of Statistical Planning and Inference 104 (2), 351–362.
- Laradji, A., 2015. Sums of totally positive functions of order 2 and applications. Statistics and Probability Letters 105, 176–180.
- Lehmann, E. L., Romano, J. P., 2005. Generalizations of the familywise error rate. Annals of Statistics 33 (3), 1138–1154.
- Li, J., Gahm, J. K., Shi, Y., Toga, A. W., 2018. Topological false discovery rates for brain mapping based on signal height. NeuroImage 167, 478–487.
- Liebscher, E., 2008. Construction of asymmetric multivariate copulas. Journal of Multivariate Analysis 99 (10), 2234–2250.
- Liebscher, E., 2011. Erratum to "Construction of asymmetric multivariate copulas" [J. Multivariate Anal. 99 (2008) 2234–2250]. Journal of Multivariate Analysis 102 (4), 869–870.
- Liu, C., Xie, J., 2014a. Large scale two sample multinomial inferences and its applications in genome-wide association studies. International Journal of Approximate Reasoning 55 (1, Part 3), 330–340.
- Liu, C., Xie, J., 2014b. Probabilistic inference for multiple testing. International Journal of Approximate Reasoning 55 (2), 654–665.

- McNeil, A. J., Nešlehová, J., 2009. Multivariate Archimedean copulas, *d*-monotone functions and  $\ell_1$ -norm symmetric distributions. Annals of Statistics 37 (5B), 3059–3097.
- Mesiar, R., Bustince, H., Fernández, J., 2010. On the α-migrativity of semicopulas, quasicopulas, and copulas. Information Sciences 180 (10), 1967–1976.
- Müller, A., Scarsini, M., 2005. Archimedean copulae and positive dependence. Journal of Multivariate Analysis 93 (2), 434–445.
- Nelsen, R. B., 2006. An Introduction to Copulas, 2nd Edition. Springer Series in Statistics. Springer, New York.
- Pawluk-Koc, M., Ziba-Palus, J., Parczewski, A., 2006. Application of false discovery rate procedure to pairwise comparisons of refractive index of glass fragments. Forensic Science International 160 (1), 53–58.
- Phillips, P. C. B., 1997. Unit-root tests. In: Kotz, S. (Ed.), Encyclopedia of Statistical Sciences, Update. Vol. 1. John Wiley & Sons, New York, pp. 531–542.
- Sarkar, S. K., 2002. Some results on false discovery rate in stepwise multiple testing procedures. Annals of Statistics 30 (1), 239–257.
- Sarkar, S. K., 2004. FDR-controlling stepwise procedures and their false negatives rates. Journal of Statistical Planning and Inference 125 (1-2), 119–137.
- Sarkar, S. K., 2008a. Generalizing Simes' test and Hochberg's stepup procedure. Annals of Statistics 36 (1), 337–363.
- Sarkar, S. K., 2008b. Rejoinder: On methods controlling the false discovery rate. Sankhyā: The Indian Journal of Statistics, Series A 70 (2), 183–185.
- Sarkar, S. K., Chang, C.-K., 1997. The Simes method for multiple hypothesis testing with positively dependent test statistics. Journal of the American Statistical Association 92 (440), 1601–1608.
- Schmidt, R., Faldum, A., Ger
  ß, J., 2015. Adaptive designs with arbitrary dependence structure based on Fisher's combination test. Statistical Methods & Applications 24 (3), 427–447.
- Schmidt, R., Faldum, A., Witt, O., Ger
  ß, J., 2014. Adaptive designs with arbitrary dependence structure. Biometrical Journal 56 (1), 86–106.
- Schweizer, B., Sklar, A., 2005. Probabilistic Metric Spaces, 2nd Edition. Dover Publications, Mineola, NY.
- Simes, R. J., 1986. An improved Bonferroni procedure for multiple tests of significance. Biometrika 73 (3), 751–754.
- Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de L'Université de Paris 8, 229–231.

- Sriboonchitta, S., Liu, J., Wiboonpongse, A., Denoeux, T., 2017. A double-copula stochastic frontier model with dependent error components and correction for sample selection. International Journal of Approximate Reasoning 80, 174–184.
- Stange, J., Bodnar, T., Dickhaus, T., 2015. Uncertainty quantification for the family-wise error rate in multivariate copula models. AStA Advances in Statistical Analysis 99 (3), 281–310.
- Storey, J. D., 2002. A direct approach to false discovery rates. Journal of the Royal Statistical Society, Series B 64 (3), 479–498.
- Wei, Z., Kim, D., 2018. On multivariate asymmetric dependence using multivariate skewnormal copula-based regression. International Journal of Approximate Reasoning 92, 376– 391.
- Westfall, P. H., Young, S. S., 1993. Resampling-Based Multiple Testing. John Wiley & Sons, New York.
- Wisadwongsa, S., Tasena, S., 2018. Bivariate quadratic copula constructions. International Journal of Approximate Reasoning 92, 1–19.
- Wu, S., 2014. Construction of asymmetric copulas and its application in two-dimensional reliability modelling. European Journal of Operational Research 238 (2), 476–485.
- Wu, W. B., 2008. On false discovery control under dependence. Annals of Statistics 36 (1), 364–380.
- Yekutieli, D., 2008. False discovery rate control for non-positively regression dependent test statistics. Journal of Statistical Planning and Inference 138 (2), 405–415.
- Yekutieli, D., Benjamini, Y., 1999. Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. Journal of Statistical Planning and Inference 82 (1-2), 171–196.
- Zimmer, D. M., Trivedi, P. K., 2006. Using trivariate copulas to model sample selection and treatment effects: application to family health care demand. Journal of Business & Economic Statistics 24 (1), 63–76.