False Discovery Rate Control under Archimedean Copula

Taras Bodnar
Department of Mathematics
Humboldt-University Berlin
Unter den Linden 6
D-10099 Berlin
Germany
e-mail: bodnar@mathematik.hu-berlin.de

and

Thorsten Dickhaus
Department of Mathematics
Humboldt-University Berlin
Unter den Linden 6
D-10099 Berlin
Germany
e-mail: dickhaus@math.hu-berlin.de

Abstract: We prove that the linear step-up procedure \( \phi_{LSU} \) considered by Benjamini and Hochberg (1995) controls the false discovery rate (FDR) in the case of dependent \( p \)-values whose dependency structure is determined by an Archimedean copula. In fact, the FDR of \( \phi_{LSU} \) is under the assumption of an Archimedean \( p \)-value copula upper-bounded by the same constant as in the case of independent \( p \)-values. Namely, the upper bound is given by \( \frac{m_0q}{m} \), where \( m \) denotes the total number of hypotheses, \( m_0 \) the number of true null hypotheses, and \( q \) the nominal FDR level. Furthermore, we establish a sharper upper bound for the FDR of \( \phi_{LSU} \) as well as a non-trivial lower bound. Application of the sharper upper bound to parametric subclasses of Archimedean \( p \)-value copulae allows us to increase the power of \( \phi_{LSU} \) by pre-estimating the copula parameter and adjusting \( q \). Based on the lower bound, a sufficient condition is obtained under which the FDR of \( \phi_{LSU} \) is exactly equal to \( \frac{m_0q}{m} \). Finally, we deal with high-dimensional multiple test problems with exchangeable test statistics by proving that the dependency structure of the corresponding vector of \( p \)-values can always be expressed by an Archimedean copula. The theoretical results are applied to important copula families, including Clayton copulae and Gumbel copulae.

AMS 2000 subject classifications: Primary 62J15, 62F05; secondary 62F03.
Keywords and phrases: Clayton copula, exchangeability, Gumbel copula, linear step-up test, multiple hypotheses testing, \( p \)-values.

1. Introduction

Control of the false discovery rate (FDR) has become a standard type I error criterion in large-scale multiple hypotheses testing. When the number \( m \)
of hypotheses to be tested simultaneously is of order $10^3 - 10^6$, as it is prevalent in many modern applications from the life sciences like genetic association analyses, gene expression studies, functional magnetic resonance imaging, or brain-computer interfacing, it is typically infeasible to model or to estimate the full joint distribution of the data. Hence, one is interested in generic procedures that control the FDR under no or only qualitative assumptions regarding this joint distribution.

The still by far most popular multiple test for FDR control, the linear step-up test $\varphi^{LSU}$ (say) considered in the seminal work by Benjamini and Hochberg (1995), operates on marginal $p$-values $p_1, \ldots, p_m$. As shown by Benjamini and Yekutieli (2001) and Sarkar (2002), $\varphi^{LSU}$ is generically FDR-controlling over the class of models that lead to positive dependency among the random $p$-values $P_1, \ldots, P_m$ in the sense of positive regression dependency on subsets (PRDS), including $p$-value distributions which are multivariate totally positive of order 2 (MTP$_2$).

In this work, we extend these findings and show that the model class over which $\varphi^{LSU}$ controls the FDR is indeed much broader and contains all models which imply a dependency structure among $P_1, \ldots, P_m$ that can be expressed by an Archimedean copula. In an asymptotic setting ($m \to \infty$), this includes all models with $p$-values or test statistics, respectively, which are exchangeable under null hypotheses, $H_0$-exchangeable for short. Such $H_0$-exchangeable test statistics occur naturally in many multiple test problems, for instance in many-to-one comparisons or if test statistics are given by jointly Studentized means (cf. Finner, Dickhaus and Roters (2007)).

While these findings are interesting in themselves, we furthermore quantify the magnitude of conservativity (non-exhaustion of the FDR level $q$) in various parametric copula models as a function of the copula parameter $\eta$. This allows a gain in power in practice by pre-estimating $\eta$ and adjusting the nominal value of $q$.

The paper is organized as follows. In Section 2, we set up the necessary notation, define our class of statistical models for $P_1, \ldots, P_m$, and recall properties and results around the FDR. Our main contributions are presented in Section 3, dealing with FDR control of $\varphi^{LSU}$ under the assumption of an Archimedean copula. Special parametric copula families are studied in Section 4, where we quantify the realized FDR of $\varphi^{LSU}$ as a function of $\eta$. Section 5 outlines methods for pre-estimation of $\eta$. We conclude with a discussion in Section 6. Lengthy proofs are deferred to Section 7.

2. Notation and preliminaries

All multiple test procedures considered in this work depend on the data only via (realized) marginal $p$-values $p_1, \ldots, p_m$ and their ordered values $p(1) \leq p(2) \leq \ldots \leq p(m)$. Hence, it suffices to model the distribution of the random vector $P = (P_1, \ldots, P_m)^T$ of $p$-values and we consider statistical models of the form $\left([0,1]^m, \mathcal{B}([0,1]^m), (\mathbb{P}_{\vartheta,\eta} : \vartheta \in \Theta, \eta \in \Xi)\right)$. In this, we assume that $\vartheta$ is the (main)
parameter of statistical interest and we identify the null hypotheses $H_i : 1 \leq i \leq m$ with non-empty subsets of $\Theta$, with corresponding alternatives $K_i = \Theta \setminus H_i$.

The null hypothesis $H_i$ is called true if $\vartheta \in H_i$ and false otherwise. We let $I_0 \equiv I_0(\vartheta) = \{1 \leq i \leq m : \vartheta \in H_i\}$ denote the index set of true hypotheses and $m_0 \equiv m_0(\vartheta) = |I_0|$ the number of true nulls. Without loss of generality, we will assume $I_0(\vartheta) = \{1, \ldots, m_0\}$ throughout the work. Analogously, we define $I = \{1, \ldots, m\}$, $I_1 \equiv I_1(\vartheta) = I \setminus I_0$ and $m_1 \equiv m_1(\vartheta) = |I_1| = m - m_0$. The intersection hypothesis $H_0 = \bigcap_{i=1}^m H_i$ will be referred to as the global (null) hypothesis.

The parameter $\eta$ is the copula parameter of the joint distribution of $P$, thus representing the dependency structure among $P_1, \ldots, P_m$. Its parameter space $\Xi$ may be of infinite dimension. In particular, in Section 3 we will consider the class of all Archimedean copulas which can be indexed by the generator function $\psi$. However, we sometimes restrict our attention to parametric subclasses, for instance the class of Clayton copulae which can be indexed by a one-dimensional copula parameter $\eta \in \mathbb{R}$. In any case, we will assume that $\eta$ is nuisance parameter in the sense that it does not depend on $\vartheta$ and that the marginal distribution of each $P_i$ is invariant with respect to $\eta$. Therefore, to simplify notation, we will write $P_{\vartheta}$ instead of $P_{\vartheta, \eta}$ if marginal $p$-value distributions are concerned. Throughout the work, the $p$-values $P_1, \ldots, P_m$ are assumed to be valid in the sense that

$$\forall 1 \leq i \leq m : \forall \vartheta \in H_i : \forall t \in [0, 1] : P_{\vartheta}(P_i \leq t) \leq t.$$  

A (non-randomized) multiple test operating on $p$-values is a measurable mapping $\varphi = (\varphi_i : 1 \leq i \leq m) : [0, 1]^m \to \{0, 1\}^m$ the components of which have the usual interpretation of a statistical test for $H_i$ versus $K_i$, $1 \leq i \leq m$. For fixed $\varphi$, we let $V_m \equiv V_m(\vartheta) = \{|i \in I_0(\vartheta) : \varphi_i = 1|\}$ denote the (random) number of false rejections (type I errors) of $\varphi$ and $R_m \equiv R_m(\vartheta) = \{|i \in \{1, \ldots, m\} : \varphi_i = 1|\}$ the total number of rejections. The FDR under $(\vartheta, \eta)$ of $\varphi$ is then defined by

$$\text{FDR}_{\vartheta, \eta}(\varphi) = \mathbb{E}_{\vartheta, \eta} \left[ \left( \frac{V_m}{R_m \vee 1} \right) \right],$$  

and $\varphi$ is said to control the FDR at level $q \in (0, 1)$ if $\sup_{\vartheta, \eta} \text{FDR}_{\vartheta, \eta}(\varphi) \leq q$. Notice that, although the trueness of the null hypotheses is determined by $\vartheta$ alone, the FDR depends on $\vartheta$ and $\eta$, because the dependency structure among the $p$-values typically influences the distribution of $\varphi$ when regarded as a statistic with values in $\{0, 1\}^m$.

The linear step-up test $\varphi^{LSU}$, also refereed to as Benjamini-Hochberg test or the FDR procedure in the literature, rejects exactly hypotheses $H_{(1)}, \ldots, H_{(k)}$, where the bracketed indices correspond to the order of the $p$-values and $k = \max\{1 \leq i \leq m : p_i(\vartheta) \leq q_i\}$ for linearly increasing critical values $q_i = iq/m$. If $k$ does not exist, no hypothesis is rejected.

The sharpest characterization of FDR control of $\varphi^{LSU}$ that we are aware of so far is given in the following theorem.
Theorem 2.1 (Finner, Dickhaus and Roters (2009)). Consider the following assumptions.

\( (D1) \forall (\vartheta, \eta) \in \Theta \times \Xi : \forall j \in I : \forall i \in I_0(\vartheta) : F_{\vartheta,\eta}(R_m \geq j | P_i \leq t) \) is non-increasing in \( t \in (0, q_j] \).

\( (D2) \forall \vartheta \in \Theta : \forall i \in I_0(\vartheta) : P_i \sim \text{UNI}([0,1]) \).

\( (I1) \forall (\vartheta, \eta) \in \Theta \times \Xi : \) The \( p \)-values \( (P_i : i \in I_0(\vartheta)) \) are independent and identically distributed (iid).

\( (I2) \forall (\vartheta, \eta) \in \Theta \times \Xi : \) The random vectors \( (P_i : i \in I_0(\vartheta)) \) and \( (P_i : i \in I_1(\vartheta)) \) are stochastically independent.

Then, the following two assertions hold true.

Under \((D1)\), \( \forall (\vartheta, \eta) \in \Theta \times \Xi : FDR_{\vartheta,\eta}(\varphi^{LSU}) \leq \frac{m_0(\vartheta)}{m} q. \) (2)

Under \((D2)-(I2)\), \( \forall (\vartheta, \eta) \in \Theta \times \Xi : FDR_{\vartheta,\eta}(\varphi^{LSU}) = \frac{m_0(\vartheta)}{m} q. \) (3)

The crucial assumption \((D1)\) is fulfilled for multivariate distributions of \( P \) which are positively regression dependent on the subset \( I_0 \) (PRDS on \( I_0 \)) in the sense of Benjamini and Yekutieli (2001). In particular, if the joint distribution of \( P \) is MTP2, then \((D1)\) holds true.

To mention also a negative result, Guo and Rao (2008) have shown that there exists a multivariate distribution of \( P \) such that the FDR of \( \varphi^{LSU} \) is equal to \( \frac{m_0 q}{m \sum_{j=1}^m j^{-1}} \), showing that \( \varphi^{LSU} \) is not generically FDR-controlling over all possible joint distributions of \( P \). The main purpose of the present work (Section 3) is to show that \( \varphi^{LSU} \) generically controls the FDR over the class of all multivariate distributions of \( P \sim P_{\vartheta,\eta} \) such that \( \Xi \ni \eta \) is the space of Archimedean copulae.

3. FDR control under Archimedean Copula

In this section, it is assumed that the joint distribution of \( P \) is given by an Archimedean copula such that

\[ F_P(p_1, ..., p_m) = F_{\vartheta,\psi}(P_1 \leq p_1, ..., P_m \leq p_m) = \psi \left( \sum_{i=1}^m \psi^{-1}(F_{P_i}(p_i)) \right), \] (4)

where the function \( \psi(\cdot) \) is the so-called copula generator and takes the role of \( \eta \) in our general setup. In (4) and throughout the work, \( F_{\xi} \) denotes the cumulative distribution function (cdf) of the variate \( \xi \). The generator \( \psi \) fully determines the type of the Archimedean copula (see, e.g. Nelsen (2006)). A necessary and sufficient condition under which a function \( \psi : \mathbb{R}_+ \to [0,1] \) with \( \psi(0) = 1 \) and \( \lim_{x \to \infty} \psi(x) = 0 \) can be used as a copula generator is that \( \psi(\cdot) \) is an \( m \)-altering function, that is, \((-1)^d \psi^{(d)}(\cdot) \geq 0\) for \( d \in \{1, 2, ..., m\} \) (cf. Müller and Scarsini (2005)).

A very useful property of the Archimedean copula is the stochastic representation of \( P \). Namely, there exists a sequence of jointly independent and identically
UNI[0,1]-distributed random variables $Y_1, \ldots, Y_m$ such that (cf. Marshall and Olkin (1988), Section 5)

$$P = (P_i : 1 \leq i \leq m) \equiv \left(F^{-1}_{P_i}\left(\psi\left(\log\left(Y_i^{-1/2}\right)\right)\right) : 1 \leq i \leq m\right),$$

where the symbol $\equiv$ denotes equality in distribution. The random variable $Z$ with Laplace transform $t \mapsto \psi(t) = \mathbb{E}[[\exp(-tZ)]]$ is independent of $Y_1, \ldots, Y_m$, and its distribution is determined by $\psi$ only. Throughout the remainder, $\mathbb{P}$ and $\mathbb{E}$ refer to the distribution of $Z$, for ease of presentation. The stochastic representation (5) shows that the type of the Archimedean copula can equivalently be expressed in terms of the random variable $Z$. Moreover, the $p$-values $(P_i : 1 \leq i \leq m)$ are conditionally independent given $Z = z$. This second property allows us to establish the following upper bound for the FDR of $\varphi^{LSU}$.

**Theorem 3.1** (Upper FDR bound). Let the copula of $P = (P_1, \ldots, P_m)^T$ be an Archimedean copula, where $P_i$ is continuously distributed on $[0,1]$ for $1 \leq i \leq m$. Then it holds that

$$\forall \vartheta \in \Theta : \forall \psi \in \Xi : \text{FDR}_{\vartheta, \psi}(\varphi^{LSU}) \leq \frac{m_0(\vartheta)}{m} q,$$

where $\Xi$ denotes the set of all generator functions of Archimedean copulæ.

Notice that this result indeed generalizes the existent findings on FDR control by Benjamini and Yekutieli (2001) and Sarkar (2002) that we have recalled in Section 1. Namely, Müller and Scarsini (2005) pointed out that an Archimedean copula possesses the MTP$_2$ property if the copula generator $\psi$ is completely monotone, i. e., if $(-1)^d\psi^{(d)}(\cdot) \geq 0$ for all $d \in \mathbb{N}$, and, hence, the FDR is controlled by $\varphi^{LSU}$ in this case. The result of Theorem 3.1 is established for the larger class of all Archimedean copulæ.

Furthermore, using equation (39) in the proof of Theorem 3.1 (see Section 7), a sharper upper bound for the FDR of $\varphi^{LSU}$ can be derived in terms of the generator $\psi$.

**Theorem 3.2** (Sharper upper bound). Let $Z$ be as in (5) and let $P^{(i)}$ consist of the $(m-1)$ remaining $p$-values obtained by dropping $P_i$ from $P$ so that $P^{(i)}_{(1)} \leq P^{(i)}_{(2)} \leq \ldots \leq P^{(i)}_{(m-1)}$. The random set $D^{(i)}_k$ is then given by

$$D^{(i)}_k = \{q_{k+1} \leq P^{(i)}_{(k)}, \ldots, q_m \leq P^{(i)}_{(m-1)}\}.$$  

For a given value $Z = z$ we define the function $T : [0,1]^m \to [0,1]^m$ by $T(p) = (T_1(p_1), \ldots, T_m(p_m))^T$ with $T_j(p_j) = \exp\left(-z\psi^{-1}(F^{-1}_{P_j}(p_j))\right)$ for $p = (p_1, \ldots, p_m)^T \in [0,1]^m$. This function transforms, for fixed $Z = z$, realizations of $P$ into realizations of $Y = (Y_1, \ldots, Y_m)^T$ given in (5). Let $D^{(i,z)}_{Y,k}$ denote the image of the set $D^{(i)}_k$ under $T$ for given $Z = z$ and let $G^{i}_{k}(z) = \mathbb{P}_{\vartheta, \psi}\left(D^{(i,z)}_{Y,k}\right)$. Then, under the assumptions of Theorem 3.1, it holds

$$\forall \vartheta \in \Theta : \text{FDR}_{\vartheta, \psi}(\varphi^{LSU}) \leq \frac{m_0(\vartheta)}{m} q - b(m, \vartheta, \psi),$$
where
\[
b(m, \vartheta, \psi) = \frac{1}{m} m \sum_{i=1}^{m_0} \sum_{k=1}^{m-1} \int_0^\infty \left( \exp \left( -\frac{z}{q_{k+1}} \right) - \exp \left( -\frac{z}{q_k} \right) \right) \times
\]
\[
( G_k(z) - G_k(z^*_k) ) dF_Z(z)
\]
\[
= \frac{1}{m} m \sum_{i=1}^{m_0} \sum_{k=1}^{m-1} \sum_{k=1}^{m-1} \frac{z^*_k}{z^*_k} \left( \exp \left( -\frac{Z}{q_{k+1}} \right) - \exp \left( -\frac{Z}{q_k} \right) \right) \times
\]
\[
(G_k(Z) - G_k(z^*_k)) \mathbb{1}_{[z^*_k, \infty)}(Z)
\]

with
\[
z^*_k = \frac{\log q_{k+1} - \log q_k}{\psi^{-1}(q_k) - \psi^{-1}(q_{k+1})} = \frac{\log (1 + 1/k)}{\psi^{-1}(1/k) - \psi^{-1}(1/(k+1))}
\]

and \(\mathbb{1}_A\) denoting the indicator function of the set \(A\).

From the practical point of view, it is problematic that \(b(m, \vartheta, \psi)\) depends on the (main) parameter \(\vartheta\) of statistical interest. In practice, one will therefore often only be able to work with \(\sup_{\vartheta \in \Theta} \{m_0(\vartheta)q/m - b(m, \vartheta, \psi)\}\). Since \(b(m, \vartheta, \psi) \geq 0\) for all \(\vartheta \in \Theta\), the latter \(\vartheta\)-free upper bound will typically still yield an improvement over the "classical" upper bound. The issue of minimization of \(b(m, \cdot, \psi)\) over \(\vartheta \in \Theta\) is closely related to the challenging task of determining least favorable parameter configurations (LFCs) for the FDR. So-called Dirac-uniform configurations are least favorable (provide upper FDR bounds) for \(\varphi_{LSU}\) under independence assumptions and are assumed to be generally least favorable for \(\varphi_{LSU}\) also in models with dependent \(p\)-values, at least for large values of \(m\) (cf., e.g., Finner, Dickhaus and Roters (2007), Blanchard et al. (2013)). Furthermore, the expectations in (8) can in general not be calculated analytically. However, they can easily be approximated by means of computer simulations. Namely, the approximation is performed by generating random numbers which behave like independent realizations of \(Z\), which completely specifies the type of the Archimedean copula, evaluating the functions \(G_k^i\) at the generated values and replacing the theoretical expectation of \(Z\) by the arithmetic mean of the resulting values of the integrand in (8). Under Dirac-uniform configurations, evaluation of \(G_k^i\) can efficiently be performed by means of recursive formulas for the joint cdf of the order statistics of \(Y\). We discuss these points in detail in Section 4.

Next, we discuss a lower bound for the FDR of \(\varphi_{LSU}\) under the assumption of an Archimedean copula.

**Theorem 3.3** (Lower FDR bound). Let the copula of \(P = (P_1, \ldots, P_m)^T\) be an Archimedean copula with generator function \(\psi\), where \(P_i\) is continuously distributed on \([0,1]\) for \(i = 1, \ldots, m\). Then it holds that
\[
\forall \vartheta \in \Theta : \text{FDR}_{\vartheta, \psi}(\varphi_{LSU}) \geq \frac{m_0 q}{m \gamma_{\min}},
\]

where
\[
γ_{\min} = \min_{\vartheta \in \Theta} \text{FDR}_{\vartheta, \psi}(\varphi_{LSU})
\]
where

\[ \gamma_{\text{min}} \equiv \gamma_{\text{min}}(\psi) = \int \min_{k \in \{1, \ldots, m\}} \left\{ \frac{\exp(-z\psi^{-1}(kq/m))}{kq/m} \right\} dF_Z(z). \] (11)

From the assertion of Theorem 3.3 we conclude that the lower bound for the FDR of \( \varphi^{LSU} \) under the assumption of an Archimedean copula crucially depends on the extreme points of the function \( g(\cdot|z) \), given by

\[ g(x|z) = \frac{\exp(-zx)}{\psi(x)} \] (12)

for \( x \in \{\psi^{-1}(q/m), \psi^{-1}(2q/m), \ldots, \psi^{-1}(q)\} \). If for all \( z > 0 \) the minimum of \( g(x|z) \) is always attained for the same index \( k^* \) (say), then \( \gamma_{\text{min}} = 1 \) and together with Theorem 3.1 we get FDR \( \varphi, \psi(\varphi^{LSU}) = m_0(\varphi)q/m \). This follows directly from the identity

\[ \int \frac{\exp(-z\psi^{-1}(k^*q/m))}{k^*q/m} dF_Z(z) = \frac{\psi(\psi^{-1}(k^*q/m))}{k^*q/m} = 1. \]

However, the latter holds true only in some specific cases. To obtain a more explicit constant \( \gamma_{\text{min}}(\psi) \) in the general case, we notice that, due to the analytic properties of \( \psi \), there exists a point \( z^* \) such that \( g(\psi^{-1}(q/m)|z) < g(\psi^{-1}(q/m)|z) \) for \( z < z^* \) and \( g(\psi^{-1}(q/m)|z) > g(\psi^{-1}(q/m)|z) \) for \( z > z^* \). The point \( z^* \) is obtained as the solution of

\[ 0 = g(\psi^{-1}(q)|z) - g(\psi^{-1}(q/m)|z) \]

\[ = \exp(-z\psi^{-1}(q)) - \exp(-z\psi^{-1}(q/m)) \]

\[ = \frac{1}{q} \left( \exp(-z\psi^{-1}(q)) - \exp\left(\log m - z\psi^{-1}\left(\frac{q}{m}\right)\right) \right) \]

\[ = \frac{\exp(-z\psi^{-1}(q))}{q} \left( 1 - \exp\left(\log m + z\left(\psi^{-1}(q) - \psi^{-1}\left(\frac{q}{m}\right)\right)\right) \right), \]

which leads to

\[ z^* = \frac{\log m}{\psi^{-1}(q/m) - \psi^{-1}(q)}. \] (13)

Next, we analyze the function \( x \mapsto g(x|z) \) for given \( z \). For its derivative with respect to \( x \), it holds that

\[ g'(x|z) = -\frac{\exp(-zx)}{(\psi(x))^2} (z\psi(x) + \psi'(x)). \]

Setting this expression to zero, we get that any extreme point of \( g(\cdot|z) \) satisfies

\[ z\psi(x) + \psi'(x) = 0. \] (14)
Let \( x_z \) be a solution of (14). Then, the second derivative of \( g(\cdot | z) \) at \( x_z \) is given by

\[
g''(x_z | z) = -\frac{\exp(-zx_z)}{(\psi(x_z))^2} \left( z\psi'(x_z) + \psi''(x_z) \right). \tag{15}\]

Substituting (14) with \( x = x_z \) in (15), we obtain

\[
g''(x_z | z) = -\frac{\exp(-zx_z)}{(\psi(x_z))^2} \left( -\frac{(\psi'(x_z))^2}{\psi(x_z)} + \psi''(x_z) \right)
= -\frac{\exp(-zx_z)}{(\psi(x_z))^3} \left( \psi(x_z)\psi''(x_z) - (\psi'(x_z))^2 \right)
\]

and application of the Cauchy-Schwarz inequality leads to

\[
\psi(x_z)\psi''(x_z) = \int \exp(-zx_z) dF_Z(z) \int z^2 \exp(-zx_z) dF_Z(z)
\geq \left( \int z \exp(-zx_z) dF_Z(z) \right)^2 = (\psi'(x_z))^2.
\]

This proves that \( g''(x_z | z) \leq 0 \) if \( x_z \) is an extreme point of \( g(x_z | z) \). Thus, any such \( x_z \) is a maximum and the minimum in (11) is attained at \( \psi^{-1}(q) \) for \( z \leq z^* \) as well as at \( \psi^{-1}(q/m) \) for \( z \geq z^* \). This allows for a more explicit characterization of the lower bound.

**Lemma 3.1.** The quantity \( \gamma_{\min} \equiv \gamma_{\min}(\psi) \) from (11) can equivalently be expressed as

\[
\gamma_{\min} = 1 - \int_0^{z^*} \left( g(\psi^{-1}(q/m) | z) - g(\psi^{-1}(q) | z) \right) dF_Z(z) \tag{16}
= 1 - E \left[ (g(\psi^{-1}(q/m) | Z) - g(\psi^{-1}(q) | Z)) 1_{[0,z^*]}(z) \right], \tag{17}
\]

where \( g(\cdot | z) \) and \( z^* \) are defined in (12) and (13), respectively.

If the integral in (16) cannot be calculated analytically, then it can easily be approximated via a Monte Carlo simulation by using the expression on the right-hand side of (17) and replacing the theoretical expectation by its pseudo-sample analogue.

Lemma 3.1 possesses several interesting applications. We consider the quantity \( \gamma_{\min} \) itself. It holds that \( 1 \geq \gamma_{\min} \geq \gamma_{\min}^* \), where

\[
\gamma_{\min} = 1 - \min \left\{ \int_0^{z^*} \sup_{z \in [0,z^*]} h(z) dF_Z(z), \int_{z^*}^{\infty} \sup_{z \in [z^*,\infty]} (-h(z)) dF_Z(z) \right\} \tag{18}
\]

with

\[
h(z) = g(\psi^{-1}(q/m) | z) - g(\psi^{-1}(q) | z)
= \frac{\exp(-z\psi^{-1}(q/m))}{q/m} - \frac{\exp(-z\psi^{-1}(q))}{q},
\]
because
\[ \int_0^{z^*} h(z) dF_Z(z) = -\int_{z^*}^{\infty} h(z) dF_Z(z). \] (19)

However, both of the integrals in (19) can be bounded by different values. To see this, we study the behavior of the function \( z \mapsto h(z) \). It holds that

\[
h'(z) = -\psi^{-1}\left(\frac{q}{m}\right) \frac{\exp(-z\psi^{-1}\left(\frac{q}{m}\right))}{q/m} + \psi^{-1}(q) \frac{\exp(-z\psi^{-1}(q))}{q}.
\]

Since \( \psi^{-1} \) is a non-increasing function, we get that there exists a unique minimum of \( h(z) \) at

\[ z^* = \frac{\log m + \log \psi^{-1}(q/m) - \log \psi^{-1}(q)}{\psi^{-1}(q/m) - \psi^{-1}(q)} \geq z^*. \]

Consequently, we get

\[
\int_0^{z^*} \sup_{z \in [0,z^*]} h(z) dF_Z(z) = \int_0^{z^*} h(0) dF_Z(z) = h(0) F_Z(z^*) = \frac{m-1}{q} F_Z(z^*),
\]

\[
\int_{z^*}^{\infty} \sup_{z \in [z^*,\infty]} (-h(z)) dF_Z(z) = \int_{z^*}^{\infty} h(z^*) dF_Z(z) = h(z^*) (1 - F_Z(z^*))
\]

\[ = \frac{\exp(-z^*\psi^{-1}(q))}{q} \left(1 - \frac{\psi^{-1}(q)}{\psi^{-1}(q/m)}\right) \times (1 - F_Z(z^*)).\]

**Corollary 3.1.** Under the assumptions of Theorem 3.3, the following two assertions hold true.

(a) If \( z^* \) from (13) does not lie in the support of \( F_Z \), i.e., if \( F_Z(z^*) = 0 \) or \( F_Z(z^*) = 1 \), then \( \gamma_{\min} = \gamma_{\min} = 1 \) and, consequently, \( FDR_{\theta,\psi}(\varphi_{LSU}) = m_0 q / m \).

(b) Assume that \( \pi_0 = \lim_{m \to \infty} m_0 / m \) exists. If \( z^* = z^*(m) \) is such that \( F_Z(z^*(m)) \to 0 \) or \( F_Z(z^*(m)) \to 1 \) as \( m \to \infty \), then

\[ \lim_{m \to \infty} FDR_{\theta,\psi}(\varphi_{LSU}) = \pi_0 q. \]

Part (b) of Corollary 3.1 motivates a deeper consideration of asymptotic or high-dimensional multiple tests, i.e., the case of \( m \to \infty \), under our general
setup. This approach has already been discussed widely in previous literature. For instance, it was called “asymptotic multiple test” by Genovese and Wasserman (2002). The case \( m \to \infty \) was also considered by Finner and Roters (1998), Storey (2002), Genovese and Wasserman (2004), Finner, Dickhaus and Roters (2007, 2009), Jin and Cai (2007), Sun and Cai (2007), and Cai and Jin (2010), among others.

Very interesting results are obtained if an infinite sequence of \( H_0 \)-exchangeable \( p \)-values is of interest. More precisely, we assume an infinite sequence \((P_i)_{i \in \mathbb{N}}\) of \( p \)-values which are absolutely continuous and uniformly distributed on \([0, 1]\) under the respective null hypothesis \( H_i \). Furthermore, we assume that \( F_1(P_1), \ldots, F_m(P_m), \ldots \) are exchangeable random variables, entailing that \( P_1, \ldots, P_m, \ldots \) themselves are exchangeable under the global hypothesis \( H_0 \). Sequences of \( H_0 \)-exchangeable \( p \)-values have already been investigated by Finner and Roters (1998) and Finner, Dickhaus and Roters (2007) in special settings. Moreover, the assumption of exchangeability is also pivotal in other areas of statistics, let us mention Bayesian analysis and validity of permutation tests. The problem of exchangeability in population genetics has been discussed by Kingman (1978).

Theorem 3.4 shows that any finite-dimensional marginal distribution of an \( H_0 \)-exchangeable \( p \)-value sequence is expressible by an Archimedean copula.

**Theorem 3.4.** Assume that the elements in the infinite sequence \((P_i)_{i \in \mathbb{N}}\) are absolutely continuous and that \((F_i(P_i))_{i \in \mathbb{N}}\) are exchangeable random variables. Then, there exists a generator \( \psi \) of an Archimedean copula such that, for any \( m \),

\[
p = (p_1, \ldots, p_m)^\top \mapsto \psi \left( \sum_{i=1}^m \psi^{-1}(p_i) \right)
\]

is a copula of \( P_1, \ldots, P_m \).

The final result of this section is an immediate consequence of Theorems 3.1 and 3.4.

**Corollary 3.2.** The linear step-up test \( \varphi^{LSU} \) controls the FDR whenever \((P_i)_{i \in \mathbb{N}}\) is an \( H_0 \)-exchangeable sequence.

### 4. Examples: Parametric copula families

In this section, we apply the theoretical results of Section 3 to several parametric families of Archimedean copulas.

#### 4.1. Independence Copula

The generator of the independence copula is given by \( \psi(t) = \exp(-t) \). Substituting \( \psi^{-1}(x) = -\ln(x) \) in (11), we get

\[
\gamma_{\text{min}} = \min_{k \in \{1, \ldots, m\}} \left\{ \frac{\exp \left( \ln \left( \frac{kq}{m} \right) \right)}{kq/m} \right\} = 1
\]
and, hence,
\[ \forall \vartheta \in \Theta : FDR_{\vartheta, \psi}(\varphi^{LSU}) = \frac{m_0(\vartheta)}{m} - q \]
under the assumption of independence. This result is in line with the previous finding reported in (3).

4.2. Clayton Copula

The generator of the Clayton copula is given by
\[ \psi(t) = (1 + \eta t)^{-1/\eta}, \quad \eta \in (0, \infty), \quad (20) \]
leading to \( \psi^{-1}(x) = (x^{-\eta} - 1) / \eta \) and to the probability density function (pdf)
\[ f_Z(z) = \frac{1}{\eta} f_{\Gamma(1/\eta, 1)}(z/\eta) = \frac{1}{\Gamma(1/\eta)} \eta^{-1/\eta} z^{1/\eta - 1} \exp(-z/\eta) \quad (21) \]
of \( Z \), where \( \Gamma \) denotes Euler’s gamma function and \( f_{\Gamma(\alpha, \beta)} \) the pdf of the gamma distribution with shape parameter \( \alpha \in (0, \infty) \) and scale parameter \( \beta \in (0, \infty) \).

For the Clayton copula, \( z^* \) is given by
\[ z^* = \frac{\log m}{\eta^{-1} (\frac{q}{m})^{-\eta} - \frac{q}{m} - \eta} = \frac{\eta \log m}{(m/q)^{\eta} - (1/q)^{\eta}}. \quad (22) \]
In Figure 1, we plot \( F_Z(z^*) \) as a function of \( \eta \) for \( m = 20 \) and \( q = 0.05 \). It is worth mentioning that the Clayton copula converges to the independence copula for \( \eta \to 0 \). In this case we get \( z^* \to 1 \) and \( f_Z(z^*) \) tends to the Dirac delta function concentrated in 1. As a result, we observe that \( F_Z(z^*) \to 1 \) as \( \eta \to 0 \) and the FDR of \( \varphi^{LSU} \) approaches \( m_0q/m \). As \( \eta \) increases, \( F_Z(z^*) \) steeply decreases and takes values very close to zero for large values of \( \eta \). Consequently, it is expected that the FDR of \( \varphi^{LSU} \) is close to \( m_0q/m \) for large values of \( \eta \), too. For \( \eta \) of moderate size, however, the FDR of \( \varphi^{LSU} \) can be much smaller than \( m_0q/m \). This is shown in Figure 2 below and discussed in detail there.

The quantity \( \gamma_{\text{min}} \) for the Clayton copula is calculated by
\[ \gamma_{\text{min}} = 1 - \eta^{-1} \int_{0}^{z^*} \frac{\exp(-z\psi^{-1}(q/m))}{q/m} f_{\Gamma(1/\eta, 1)}(z/\eta) dz + \eta^{-1} \int_{0}^{z^*} \frac{\exp(-z\psi^{-1}(1))}{q} f_{\Gamma(1/\eta, 1)}(z/\eta) dz \]
\[ = 1 - I_1^C + I_2^C, \quad (23) \]
Fig 1. The value $F_Z(z^*)$ as a function of $\eta$ for $m = 20$ and $q = 0.05$ under the assumption of a Clayton copula.
where

\[
I_1^C = \int_0^z z^{1/\eta - 1} \exp \left( -\frac{z}{\eta} \left( \frac{m}{q} \right)^\eta - 1 \right) dz
\]

and, similarly,

\[
I_2^C = \int_0^z z^{1/\eta - 1} \exp \left( -\frac{z}{\eta} \left( \frac{m}{q} \right)^\eta \right) dz
\]

Hence, from Theorem 3.3 we get for all \( \vartheta \in \Theta \) that

\[
FDR_{\vartheta, \eta}(\varphi_{LSU}) \geq \frac{m_0 q}{m} \left( 1 + \frac{\ln m}{m^\eta - 1} \right) - \frac{F_{\Gamma(1/\eta, 1)} \left( \ln m / m^\eta - 1 \right)}{F_{\Gamma(1/\eta, 1)} \left( \ln m / m^\eta - 1 \right)}.
\]

Next, we consider the sharper upper bound for the FDR in the case of Clayton copulae in detail. As outlined in the discussion around Theorem 3.2, a so-called Dirac-uniform configuration (cf., e.g., Blanchard et al. (2013) and references therein) is assumed for \( P \) in case of \( m_0 < m \). Namely, the p-values \( (P_i : i \in I_1(\vartheta)) \) are assumed to be \( P_\vartheta \)-almost surely equal to 0. Under assumptions (I1)-(I2) from Theorem 2.1, Dirac-uniform configurations are least favorable (provide upper bounds) for the FDR of \( \varphi_{LSU} \), see Benjamini and Yekutieli (2001). In the case of dependent p-values, such general results are yet lacking, but it is assumed that Dirac-uniform configurations yield upper FDR bounds for \( \varphi_{LSU} \) also under dependency, at least for large \( m \) (cf., e.g., Finner, Dickhaus and Roters (2007)).

Under a Dirac-uniform configuration, the sharper upper bound for the FDR of \( \varphi_{LSU} \) is expressed by (see Theorem 3.2)

\[
b(m, \vartheta, \eta) = \frac{q}{m} \sum_{i=1}^{m_0} \sum_{k=m_1+1}^{m-1} E \left[ \left( \exp \left( -Z\psi^{-1}(q_{k+1}) / q_{k+1} \right) - \exp \left( -Z\psi^{-1}(q_k) / q_k \right) \right) \right]
\times \left( G_k^i(Z) - G_k^i(z_k^*) \right) \mathbb{1}_{[z_k^*, \infty)}(Z)
\]

where \( z_k^* \) is given in (9) and the random set \( D_{Y,k}^{(i,z)} \) the probability of which is given by \( G_k^i(z) \) can under Dirac-uniform configurations equivalently be expressed as

\[
D_{Y,k}^{(i,z)} = \left\{ \exp \left( -z\psi^{-1}(q_{k+1}) \right) \leq Y_{(k)}^{(i)}, ..., \exp \left( -z\psi^{-1}(q_m) \right) \leq Y_{(m-1)}^{(i)} \right\}.
\]
The last equality follows from the fact that \( Y_{(k)}^{(i)}, \ldots, Y_{(m-1)}^{(i)} \) almost surely correspond to \( p \)-values associated with true null hypotheses, i.e.,

\[
F_{p_{(k)}^{(i)}}(x) = \ldots = F_{p_{(m-1)}^{(i)}}(x) = x.
\]

Moreover, since each of the \( Y_{(k)}^{(i)}, \ldots, Y_{(m-1)}^{(i)} \) is obtained by the same isotonic transformation from the corresponding element in the sequence \( P_{(k)}^{(i)}, \ldots, P_{(m-1)}^{(i)} \), we get that \( Y_{(k)}^{(i)}, \ldots, Y_{(m-1)}^{(i)} \) is an increasing sequence of independent and identically \( \text{UNI}[0,1] \)-distributed random variables. Hence, the probabilities \( G_k(z) = \mathbb{P}_{\theta,\eta}(D_{Y,k}^{(i,z)}) \) for \( k \in \{m_1 + 1, \ldots, m - 1\} \) can be calculated recursively, for instance by making use of Bolshev's recursion (see, e.g., Shorack and Wellner (1986), p. 366).

In general, Bolshev's recursion is defined in the following way. Let \( 0 \leq a_1 \leq a_2 \leq \ldots \leq a_n \leq 1 \) be real constants and let \( U_{(1)} \leq U_{(2)} \leq \ldots \leq U_{(n)} \) be the order statistics of independent and identically \( \text{UNI}[0,1] \)-distributed random variables. We let \( \hat{P}_n(a_1, \ldots, a_n) = P(a_1 \leq U_{(1)}, \ldots, a_n \leq U_{(n)}) \). Then, the probability \( \hat{P}_n(a_1, \ldots, a_n) \) is calculated recursively by

\[
\hat{P}_n(a_1, \ldots, a_n) = 1 - \sum_{j=1}^{n} \binom{n}{j} a_j \hat{P}_{n-j}(a_{j+1}, \ldots, a_n). \tag{25}
\]

Application of (25) with \( n = m_0 - 1 \) and

\[
a_j = \begin{cases} 
0 & \text{for } j \in \{1, \ldots, k - m_1 - 1\} \\
\exp \left( -2z^{-1} (q_{j+m_1+1}) \right) & \text{for } j \in \{k - m_1, \ldots, m_0 - 1\}
\end{cases}
\]

for \( k \in \{m_1 + 1, \ldots, m - 1\} \) as well as numerical integration with respect to the distribution of \( Z \) over \([z_k^+, \infty)\) lead to a numerical approximation of the sharper upper bound for the FDR of \( \varphi^{LSU} \) under Dirac-uniform configurations.

In Figure 2 we present the lower bound (dashed red line), the upper bound (dashed blue line), the sharper upper bound (solid green line), and the simulated values of the FDR of \( \varphi^{LSU} \) (solid black line) as a function of the parameter \( \eta \) of a Clayton copula. We put \( m = 20, q = 0.05 \), and \( m_0 = 16 \). The \( p \)-values which correspond to the false null hypotheses have been set to zero. The simulated values are obtained by using \( 10^5 \) independent repetitions. We observe that the FDR of \( \varphi^{LSU} \) starts at \( m_0 q/m = 0.04 \) for \( \eta = 0 \) and decreases to a minimum of approximately 0.023 at \( \eta \approx 1.7 \). This value is much smaller than the nominal level \( q \), offering some room for improvement of \( \varphi^{LSU} \) for a broad range of values of \( \eta \). After reaching its minimum, the FDR of \( \varphi^{LSU} \) increases and tends to 0.04 as \( \eta \) increases. This behavior of the FDR of \( \varphi^{LSU} \) is as expected from the values of \( F_Z(z^*) \), as discussed around Figure 1.

In contrast to the "classical" upper bound, the sharper upper bound reproduces the behavior of the simulated FDR values very well. It provides a good approximation of the true values of the FDR of \( \varphi^{LSU} \) for all considered values
of $\eta$. In particular, it is much smaller than the "classical" upper bound for moderate values of $\eta$. Consequently, application of the sharper upper bound can be used to improve the power of the multiple testing procedure by adjusting the nominal value of $q$ depending on $\eta$. If $\eta$ is unknown, we propose techniques for pre-estimating it in Section 5. Finally, we note that the lower bound seems not to be informative. It is close to zero even for moderate values of $\eta$.

Fig 2. Lower bound (dashed red line), upper bound (dashed blue line), sharper upper bound (solid green line), and simulated values of the FDR of $\phi^{LSU}$ (solid black line) as a function of $\eta$ for Clayton copulae. We put $m = 20$, $q = 0.05$, and $m_0 = 16$. Simulated values were obtained by using $10^5$ independent pseudo realizations of $Z$.

4.3. Gumbel Copula

The generator of the Gumbel copula is given by

$$\psi(x) = \exp\left(-x^{1/\eta}\right), \quad \eta \geq 1,$$

which leads to $\psi^{-1}(x) = (-\ln x)^\eta$ and a stochastic representation

$$Z \overset{d}{=} \left(\cos\left(\frac{\pi}{2\eta}\right)\right)^\eta Z_0, \quad \eta > 1,$$ (27)
for $Z$, where the random variable $Z_0$ has a stable distribution with index of stability $1/\eta$ and unit skewness. The cdf of $Z_0$ is given by (cf. Chambers, Mallows and Stuck (1976), p. 341)

$$F_{Z_0}(z) = \frac{1}{\pi} \int_0^\pi \exp \left(-z^{-1/(\eta-1)}a(v)\right) dv \quad \text{with}$$

$$a(v) = \frac{\sin((1-\eta)v/\eta)(\sin(v/\eta))^{1/(\eta-1)}}{(\sin v)^{\eta/(\eta-1)}}, \quad v \in (0, \pi).$$

Although (27) in connection with $F_{Z_0}$ characterizes the distribution of $Z$ completely, the integral representation of $F_{Z_0}$ may induce numerical issues with respect to implementation. Somewhat more convenient from this perspective is the following result. Namely, Kanter (1975) obtained a stochastic representation of $Z_0$, given by

$$Z_0 = \left(\frac{a(U)}{W}\right)^{\eta-1}, \quad (28)$$

where $U$ and $W$ are stochastically independent, $W$ is standard exponentially distributed and $U \sim \text{UNI}(0, \pi)$. We used (28) for simulating $Z_0$ and, consequently, $Z$.

For the Gumbel copula we get

$$z^* = \ln m \left(-\ln \frac{q}{m}\right)^{\eta} - \left(-\ln \frac{q}{m}\right)^{\eta} = \ln \frac{m}{q} - \left(\ln \frac{1}{q}\right)^{\eta}. \quad (29)$$

In Figure 3, we plot $F_Z(z^*)$ as a function of $\eta$ for $m = 20$ and $q = 0.05$. A similar behavior as in the case of the Clayton copula is present. If $\eta = 1$ then the Gumbel copula coincides with the independence copula. Hence, $F_Z(z^*) = 1$ and, consequently, the FDR of $\varphi^{LSU}$ is equal to $mq/m$ in this case. As $\eta$ increases, $F_Z(z^*)$ decreases and it approaches 0 for larger values of $\eta$. Hence, FDR$_{\varphi^{LSU}}(\varphi^{LSU})$ tends to $mq/m$ as $\eta$ becomes considerably large. For moderate values of $\eta$, FDR$_{\varphi^{LSU}}(\varphi^{LSU})$ can again be much smaller than $mq/m$, in analogy to the situation in models with Clayton copulae.

Recall from (17) that

$$\gamma_{\min} = 1 - \mathbb{E} \left[g_1(Z) \mathbf{1}_{[0,z^*]}(Z)\right], \quad (30)$$

where $g_1(Z) = g \left(\psi^{-1}(q/m)|Z\right) - g \left(\psi^{-1}(q)|Z\right)$. For the Gumbel copula, we obtain

$$g_1(Z) = \exp \left(-Z\psi^{-1}(q/m)\right) - \exp \left(-Z\psi^{-1}(q)\right)$$

$$= \frac{\exp \left(-Z \left(\ln \frac{m}{q}\right)^{\eta}\right)}{q/m} - \frac{\exp \left(-Z \left(\ln \frac{1}{q}\right)^{\eta}\right)}{q}.$$
Fig 3. The value $F_{Z}(z^{*})$ as a function of $\eta$ for $m = 20$ and $q = 0.05$ under the assumption of a Gumbel copula. The graph was obtained via simulations by generating $10^6$ independent pseudo realizations of $Z$ according to (27) and (28).
\[ \frac{mqq}{m} (1 - \mathbb{E}[g_1(Z)1_{[0,z^*]}(Z)]) \leq \text{FDR}_{\varphi,\eta}(\varphi^{LSU}) \leq \frac{mqq}{m}. \] (31)

The sharper upper bound from Theorem 3.2 can be calculated by using Bolshov’s recursion similarly to the discussion around (25), but here with \( \psi \) as in (26). Figure 4 displays the lower bound (dashed red line), the upper bound (dashed blue line), the sharper upper bound (solid green line), and simulated values of \( \text{FDR}_{\varphi,\eta}(\varphi^{LSU}) \) (solid black line) as functions of \( \eta \). Again, we chose \( m = 20, q = 0.05, \) and \( m_0 = 16 \). The \( p \)-values corresponding to the false null hypotheses were all set to zero, as in the case of Clayton copulae. The simulated values were obtained by generating \( 10^5 \) independent pseudo realizations of \( Z \).

Similarly to the case of the Clayton copula, the curve of simulated FDR values has a \( U \)-shape. It starts at \( m_0q/m = 0.04 \) and drops to its minimum of approximately 0.024 for values of \( \eta \) around 6.6. For such values of \( \eta \), the black curve is considerably below the classical upper bound of 0.04. In contrast, the sharper upper bound gives a much tighter approximation of the simulated FDR values in such cases and reproduces the \( U \)-shape over the entire range of values for the parameter \( \eta \) of the Gumbel copula. As a result, its application can be used to improve power by adjusting the nominal value of \( q \) and thereby increasing the probability to detect false null hypotheses. The lower bound from (31) (corresponding to the dashed red curve) has been obtained by approximating the expectation in (30) via simulations. As in the case of the Clayton copula, the lower bound is not too informative for the model class that we have considered here (Dirac-uniform configurations).

5. Empirical copula calibration

In the previous section we studied the influence of the copula parameter \( \eta \) on the FDR of \( \varphi^{LSU} \) under several parametric families of Archimedean copulae. It turned out that adapting \( \varphi^{LSU} \) to the degree of dependency in the data by adjusting the nominal value of \( q \) based on the sharper upper bound from Theorem 3.2 is a promising idea, because the unadjusted procedure may lead to a considerable non-exhaustion of \( q \), cf. Figures 2 and 4. Due to the decision rule of a step-up test, this also entails suboptimal power properties of \( \varphi^{LSU} \) when applied ”as is” to models with Archimedean \( p \)-value copulae.

In practice, however, often the copula parameter itself is an unknown quantity. Hence, the outlined adaptation of \( q \) typically requires some kind of pre-estimation of \( \eta \) before multiple testing is performed. Although this is not in the main focus of the present work, we therefore outline possibilities for estimating \( \eta \) and for quantifying the uncertainty of the estimation in this section.

One class of procedures relies on resampling, namely via the parametric bootstrap or via permutation techniques if \( H_1, \ldots, H_m \) correspond to marginal two-sample problems. Pollard and van der Laan (2004) provided an extensive comparison of both approaches and argued that the permutation method reproduces the correct null distribution only under some conditions. However, if these conditions are met, the permutation approach is often superior to bootstrapping.
Fig 4. Lower bound (dashed red line), upper bound (dashed blue line), the sharper upper bound (solid green line), and simulated values of the FDR of $\varphi_{LSU}$ (solid black line) as a function of the parameter $\eta$ of a Gumbel copula. We put $m = 20$, $q = 0.05$, and $m_0 = 16$. Simulated values are based on on $10^5$ independent pseudo realizations of $Z$. 

FDR, Gumbel Copula

\begin{center}
\begin{tikzpicture}
\begin{axis}[
width=\textwidth,
height=0.4\textwidth,
axis lines=left,
axis line style={-},
axis y line style={-},
axis x line style={-},
xlabel=$\eta$,
ylabel=FDR,
xtick={2,4,6,8,10,12,14,16,18,20},
xticklabels={2,4,6,8,10,12,14,16,18,20},
ytick={0.01,0.02,0.03,0.04},

\addplot[red,dashed] table[header=first row=true] {data1.dat};
\addlegendentry{Upper Bound}
\addplot[green,dashed] table[header=first row=true] {data2.dat};
\addlegendentry{Sharper Upper Bound}
\addplot[red] table[header=first row=true] {data3.dat};
\addlegendentry{Lower Bound}
\addplot[black] table[header=first row=true] {data4.dat};
\addlegendentry{Simulated Values}
\end{axis}
\end{tikzpicture}
\end{center}
(see also Westfall and Young (1993) and Meinshausen, Maathuis and Bühlmann (2011)). Furthermore, it is essential to keep in mind that both bootstrap and permutation-based methods estimate the distribution of the vector $\mathbf{p}$ under the global null hypothesis $H_0$. Hence, the assumption that $\eta$ does not depend on $\vartheta$ is an essential prerequisite for the applicability of such resampling methods for estimating $\eta$. Notice that the latter assumption is an informal description of the "subset pivotality" condition introduced by Westfall and Young (1993). The resampling methods developed by Dudoit and van der Laan (2008) can dispense with subset pivotality in special model classes, but for the particular task of estimating the copula parameter this assumption seems indispensable.

Estimation of $\eta$ and uncertainty quantification of the estimation based on resampling is generally performed by applying a suitable estimator $\hat{\eta}$ to the re-(pseudo) samples. In the context of Archimedean copulae the two most widely applied estimation procedures are the maximum likelihood method (see, e.g. Joe (2005), Hofert, Mächler and McNeil (2012)) and the method of moments (referred to as "realized copula" approach by Fengler and Okhrin (2012)).

Hofert, Mächler and McNeil (2012) considered the estimation of the parameter of an Archimedean copula with known margins by the maximum likelihood approach. To this end, they derived analytic expressions for the derivatives of the copula generator for several families of Archimedean copulae, as well as formulas for the corresponding score functions. Using these results and assuming a regular model, an elliptical asymptotic confidence region for the copula parameter $\eta$ can be obtained by applying general limit theorems for maximum likelihood estimators (see Hofert, Mächler and McNeil (2012) for details and the calculations for different types of Archimedean copulae).

In the context of the method of moments, Kendall’s tau is often considered. For a bivariate Archimedean copula with generator $\psi$ of marginally UNI[0,1]-distributed variates $P_1$ and $P_2$, it is given by

$$\tau_{P_1,P_2} = 4 \int_0^1 \int_0^1 F_{(P_1,P_2)}(u,v)dF_{(P_1,P_2)}(u,v) - 1$$

$$= 1 - 4 \int_0^{\psi^{-1}(0)} t|\psi'(t)|^2 dt,$$

(32)

cf. McNeil and Neslehová (2009).

The right-hand side of (32) can analytically be calculated for some families of Archimedean copulae. For instance, for a Clayton copula with parameter $\eta$ it is given by $\tau(\eta) = \eta/(2 + \eta)$, while it is equal to $\tau(\eta) = (\eta - 1)/\eta$ for a Gumbel copula with parameter $\eta$ (see Nelsen (2006), p. 163-164). Based on such moment equations, Fengler and Okhrin (2012) suggested the "realized copula" method for empirical calibration of a one-dimensional parameter $\eta$ of an $m$-variate Archimedean copula. The method considers all $m(m-1)/2$ distinct pairs of the $m$ underlying random variables, replaces the population versions of $\tau(\eta)$ by the corresponding sample analogues, and finally aggregates the resulting $m(m-1)/2$ estimates in an appropriate manner. More specifically, consider the functions $g_{ij}(\eta) = \hat{\tau}_{ij} - \tau(\eta)$ for $1 \leq i < j \leq m$ and define $q(\eta) = (g_{ij}(\eta) : \quad \text{(32)}}$
1 ≤ i < j ≤ m)⊤, where  \( \hat{\tau}_{ij} \) is the sample estimator of Kendall’s tau (see, e. g., Nelsen (2006), Section 5.1.1). The resulting estimator for  \( \eta \) is then obtained by
\[
\hat{\eta} = \arg\min_{\eta} \left\{ q(\eta)^{\top} W q(\eta) \right\}
\]
for an appropriate weight matrix  \( W \in \mathbb{R}^{\binom{m}{2} \times \binom{m}{2}} \). An application of the realized copula method to resampled  \( p \)-values generated by permutations in the context of multiple testing for differential gene expression has been demonstrated by Dickhaus and Gierl (2013). Multivariate extensions of Kendall’s tau and central limit theorems for the sample versions have been derived by Genest, Nešlehová and Ben Ghorbal (2011). These results can be used for uncertainty quantification of the moment estimation of  \( \eta \) by constructing asymptotic confidence regions.

6. Discussion

We have shown that the class of models over which  \( \varphi^{LSU} \) controls the FDR is considerably broader than the class of PRDS models and in particular contains all models entailing an Archimedean  \( p \)-value copula. Since models with  \( H_0 \)-exchangeable  \( p \)-values fall into this class at least asymptotically (see Theorem 3.4), our findings extend those of Finner, Dickhaus and Roters (2007) who investigated infinite sequences of  \( H_0 \)-exchangeable  \( p \)-values in Gaussian models. While our general results in Section 3 qualitatively extend the theory (in the sense that  \( \varphi^{LSU} \) controls the FDR over the broader class), our results in Section 4 regarding Clayton and Gumbel copulae are quantitatively very much in line with the findings for Gaussian and  \( t \)-copulae reported by Finner, Dickhaus and Roters (2007). Namely, over a broad class of models with dependent  \( p \)-values, the FDR of  \( \varphi^{LSU} \) as a function of the dependency parameter has a  \( U \)-shape and becomes smallest for medium strength of dependency among the  \( p \)-values. This behavior can be exploited by adjusting  \( q \) in order to adapt to  \( \eta \). We have presented an explicit adaptation scheme based on the upper bound from Theorem 3.2. To the best of our knowledge, this kind of adaptation is novel to FDR theory.

It is beyond the scope of the present work to investigate which parametric class of copulae is appropriate for which kind of real-life application. Relatedly, the problem of model misspecification (i. e., quantification of the approximation error if the true model does not belong to the class with Archimedean  \( p \)-value copulae and is approximated by the (in some suitable norm) closest member of this class) could not be addressed here, but is a challenging topic for future research. One particularly interesting issue in this direction is FDR control for finite sequences of  \( H_0 \)-exchangeable  \( p \)-values.

Finally, we would like to mention that the empirical variance of the false discovery proportion  \( \text{FDP}_{\varphi^{LSU}}(\varphi^{LSU}) = V_m / \max(R_m, 1) \) was large in all our simulations, implying that the random variable  \( \text{FDP}_{\varphi^{LSU}}(\varphi^{LSU}) \) was not well concentrated around its expected value  \( \text{FDR}_{\varphi^{LSU}}(\varphi^{LSU}) \). This is a known effect for models with dependent  \( p \)-values (see, e. g., Finner, Dickhaus and Roters...
(2007), Delattre and Roquain (2011), Blanchard et al. (2013)) and provokes the question if FDR control is a suitable criterion under dependency at all. Maybe more stringent in dependent models is control of the false discovery exceedance rate, meaning to design a multiple test $\varphi$ ensuring that $\text{FDX}_{\varphi}(\varphi) = \mathbb{P}_{\varphi}(\text{FDP}_{\varphi}(\varphi) > c) \leq \gamma$, for user-defined parameters $c$ and $\gamma$. In any case, practitioners should be (made) aware of the fact that controlling the FDR with $\varphi_{LSU}$ does not necessarily imply that the FDP for their particular experiment is small, at least if dependencies among $P_1, \ldots, P_m$ have to be assumed as it is typically the case in applications.

7. Proofs

**Proof of Theorem 3.1**

Following Benjamini and Yekutieli (2001), an analytic expression for the FDR of $\varphi_{LSU}$ is given by

$$\text{FDR}_{\varphi_{LSU}} = \sum_{i=1}^{m_0} \sum_{k=1}^{m} \frac{1}{k} \mathbb{P}_{\vartheta,\eta}(A^{(i)}_k),$$

where $A^{(i)}_k = \{P_i \leq q_k \cap C^{(i)}_k\}$ denotes the event that $k$ hypotheses are rejected one of which is $H_i$ (a true null hypothesis) and $C^{(i)}_k$ is the event that $k - 1$ hypotheses additionally to $H_i$ are rejected. It holds that $(C^{(i)}_k : 1 \leq k \leq m)$ are disjoint and that $\bigcup_{k=1}^{m} C^{(i)}_k = [0, 1]^{m-1}$.

Let $D^{(i)}_k = \bigcup_{j=1}^{k} C^{(i)}_j$ for $k = 1, \ldots, m$ denote the event that the number of rejected null hypotheses is at most $k$. In terms of $\mathbb{P}^{(i)}$ introduced in Theorem 3.2, the random set $D^{(i)}_k$ is given by

$$D^{(i)}_k = \{q_{k+1} \leq P^{(i)}_{(k)} \leq \cdots \leq q_m \leq P^{(i)}_{(m-1)}\}.$$  

Next, we prove that

$$\frac{\mathbb{P}_{\vartheta,\eta}(P_i \leq q_k \cap D^{(i)}_k)}{q_k} \leq \frac{\mathbb{P}_{\vartheta,\eta}(P_i \leq q_{k+1} \cap D^{(i)}_k)}{q_{k+1}}.\quad (36)$$

To this end, we consider the function $T$ introduced in Theorem 3.2, which transforms a possible realization of the original $p$-values $\mathbb{P}$ into a realization of $\mathbb{Y}$ for $Z = z$, where $\mathbb{Y} = (Y_1, \ldots, Y_m)^\top$ and $Z$ are as in (5). Because each component of this multivariate transformation is a monotonically increasing function which fully covers the interval $[0, 1]$, the resulting transformation bijectively transforms the set $[0, 1]^m$ into itself. Let $C^{(i,z)}_{Y:k}$ and $D^{(i,z)}_{Y:k}$ denote the images of the sets $C^{(i)}_k$ and $D^{(i)}_k$ under $T$ for given $Z = z$. Then

(a) $C^{(i,z)}_{Y:k}$ are disjoint, i.e., $C^{(i,z)}_{Y:k_1} \cap C^{(i,z)}_{Y:k_2} = \emptyset$ for $1 \leq k_1 \neq k_2 \leq m,$
(b) \( D^{(i,z)}_{Y,k} = \bigcup_{j=1}^{k} C^{(i,z)}_{Y,j} \),
(c) \( D^{(i,z)}_{Y,m} = \bigcup_{j=1}^{m} C^{(i,z)}_{Y,j} = [0,1]^{m-1} \).

Statements (a) - (c) follow directly from the facts that each \( T_j \) is a monotonically increasing function and \( T \) is a one-to-one transformation with image equal to \([0,1]^{m}\). Moreover, we obtain

\[
D^{(i,z)}_{Y,k} = \{ \forall k \leq j \leq m-1 : Y^{(i)}_{(j)} \geq \exp \left( -z\psi^{-1} \left( F_{\psi^{-1}}^{(i)}(q_{j+1}) \right) \right) \},
\]

(37)

where \( Y^{(i)} \) is the \((m-1)\)-dimensional vector obtained from \( Y = (Y_1, \ldots, Y_m)^T \) by deleting \( Y_i \). The last equality shows that \( D^{(i,z)}_{Y,k} \subseteq D^{(i,z)}_{Y,k} \) for \( z_1 \leq z_2 \) and, hence, that \( G^i_k \), given by \( G^i_k(z) = \mathbb{P}_{\vartheta,\eta}(D^{(i,z)}_{Y,k}) \), is an increasing function in \( z \).

Returning to (36), we obtain

\[
\begin{align*}
\mathbb{P}_{\vartheta,\eta} \left( P_i \leq q_{k+1} \cap D^{(i)}_{k} \right) & - \mathbb{P}_{\vartheta,\eta} \left( P_i \leq q_k \cap D^{(i)}_{k} \right) \\
& = \int \left( \frac{\mathbb{P}_{\vartheta,\eta} \left( P_i \leq q_{k+1} \cap D^{(i)}_{k} | Z = z \right)}{q_{k+1}} - \frac{\mathbb{P}_{\vartheta,\eta} \left( P_i \leq q_k \cap D^{(i)}_{k} | Z = z \right)}{q_k} \right) dF_Z(z) \\
& = \int \left( \frac{\mathbb{P}_{\vartheta,\eta} \left( P_i \leq q_{k+1} | Z = z \right) \mathbb{P}_{\vartheta,\eta} \left( D^{(i)}_{k} | Z = z \right)}{q_{k+1}} - \frac{\mathbb{P}_{\vartheta,\eta} \left( P_i \leq q_k | Z = z \right) \mathbb{P}_{\vartheta,\eta} \left( D^{(i)}_{k} | Z = z \right)}{q_k} \right) dF_Z(z) \\
& = \int \left( \frac{\mathbb{P}_{\vartheta,\eta} \left( Y_i \leq \exp \left( -z\psi^{-1}(q_{k+1}) \right) \right)}{q_{k+1}} - \frac{\mathbb{P}_{\vartheta,\eta} \left( Y_i \leq \exp \left( -z\psi^{-1}(q_k) \right) \right)}{q_k} \right) \mathbb{P}_{\vartheta,\eta} \left( D^{(i,z)}_{Y,k} \right) dF_Z(z) \\
& = \int \left( \frac{\exp \left( -z\psi^{-1}(q_{k+1}) \right)}{q_{k+1}} - \frac{\exp \left( -z\psi^{-1}(q_k) \right)}{q_k} \right) G^i_k(z) dF_Z(z).
\end{align*}
\]

(38)
Next, we analyze the difference under the last integral. It holds that
\[
\frac{\exp(-z\psi^{-1}(q_{k+1}))}{q_{k+1}} - \frac{\exp(-z\psi^{-1}(q_k))}{q_k} = \exp(-\log q_{k+1} - z\psi^{-1}(q_{k+1})) - \exp(-\log q_k - z\psi^{-1}(q_k)) = \exp(-\log q_k - z\psi^{-1}(q_k)) \times \left(\exp(-\log q_{k+1} + \log q_k - z\psi^{-1}(q_{k+1}) + z\psi^{-1}(q_k)) - 1\right).
\]
The last expression is nonnegative if and only if
\[-\log q_{k+1} + \log q_k - z\psi^{-1}(q_{k+1}) + z\psi^{-1}(q_k) \geq 0.
\]
Hence, for \(z \geq z^*_k\) with \(z^*_k\) given in (9), the function under the integral in (38) is positive and for \(z \leq z^*_k\) it is negative. Application of this result leads to
\[
\mathbb{P}_{\theta, \eta}(P_i \leq q_{k+1} \cap D_k^{(i)}) - \mathbb{P}_{\theta, \eta}(P_i \leq q_k \cap D_k^{(i)}) = \int_0^{z_k} \left(\frac{\exp(-z\psi^{-1}(q_{k+1}))}{q_{k+1}} - \frac{\exp(-z\psi^{-1}(q_k))}{q_k}\right) G_k^i(z) dF_Z(z)
+ \int_{z_k}^{\infty} \left(\frac{\exp(-z\psi^{-1}(q_{k+1}))}{q_{k+1}} - \frac{\exp(-z\psi^{-1}(q_k))}{q_k}\right) G_k^i(z) dF_Z(z)
\geq G_k^i(z_k) \int_0^{z_k} \left(\frac{\exp(-z\psi^{-1}(q_{k+1}))}{q_{k+1}} - \frac{\exp(-z\psi^{-1}(q_k))}{q_k}\right) dF_Z(z)
+ \int_{z_k}^{\infty} \left(\frac{\exp(-z\psi^{-1}(q_{k+1}))}{q_{k+1}} - \frac{\exp(-z\psi^{-1}(q_k))}{q_k}\right) G_k^i(z) dF_Z(z).
\]
Because of
\[
\int \left(\frac{\exp(-z\psi^{-1}(q_{k+1}))}{q_{k+1}} - \frac{\exp(-z\psi^{-1}(q_k))}{q_k}\right) dF_Z(z) = \int \frac{\exp(-z\psi^{-1}(q_{k+1}))}{q_{k+1}} dF_Z(z) - \int \frac{\exp(-z\psi^{-1}(q_k))}{q_k} dF_Z(z)
= \frac{\psi\left(\psi^{-1}(q_{k+1})\right)}{q_{k+1}} - \frac{\psi\left(\psi^{-1}(q_k)\right)}{q_k} = \frac{q_{k+1}}{q_{k+1}} - \frac{q_k}{q_k} = 0
\]
we get
\[
\int_0^{z_k} \left(\frac{\exp(-z\psi^{-1}(q_{k+1}))}{q_{k+1}} - \frac{\exp(-z\psi^{-1}(q_k))}{q_k}\right) dF_Z(z) = - \int_{z_k}^{\infty} \left(\frac{\exp(-z\psi^{-1}(q_{k+1}))}{q_{k+1}} - \frac{\exp(-z\psi^{-1}(q_k))}{q_k}\right) dF_Z(z).
\]
and, consequently,

\[
\begin{align*}
&\frac{\mathbb{P}_{\vartheta, \eta} \left( P_i \leq q_k \cap D_k^{(i)} \right)}{q_{k+1}} - \frac{\mathbb{P}_{\vartheta, \eta} \left( P_i \leq q_k \cap D_k^{(i)} \right)}{q_k} \\
\geq & -G_k(z_k^*) \int_{z_k^*}^{\infty} \left( \frac{\exp \left(-z\psi^{-1}(q_{k+1})\right)}{q_{k+1}} - \frac{\exp \left(-z\psi^{-1}(q_k)\right)}{q_k} \right) dF_Z(z) \\
&+ \int_{z_k^*}^{\infty} \left( \frac{\exp \left(-z\psi^{-1}(q_{k+1})\right)}{q_{k+1}} - \frac{\exp \left(-z\psi^{-1}(q_k)\right)}{q_k} \right) G_k(z) dF_Z(z) \\
&= \int_{z_k^*}^{\infty} \left( \frac{\exp \left(-z\psi^{-1}(q_{k+1})\right)}{q_{k+1}} - \frac{\exp \left(-z\psi^{-1}(q_k)\right)}{q_k} \right) (G_k(z) - G_k(z_k^*)) dF_Z(z) \\
\geq & 0,
\end{align*}
\]

since both the differences under the integral in (39) are positive. This completes the proof of (36).

Using (36), we get for all \(1 \leq k \leq m-1\) that

\[
\begin{align*}
&\frac{\mathbb{P}_{\vartheta, \eta} \left( P_i \leq q_k \cap D_k^{(i)} \right)}{q_{k+1}} + \frac{\mathbb{P}_{\vartheta, \eta} \left( P_i \leq q_k \cap D_k^{(i)} \right)}{q_k} \\
\leq & \frac{\mathbb{P}_{\vartheta, \eta} \left( P_i \leq q_{k+1} \cap D_k^{(i)} \right)}{q_{k+1}} + \frac{\mathbb{P}_{\vartheta, \eta} \left( P_i \leq q_{k+1} \cap C_{k+1}^{(i)} \right)}{q_{k+1}} \\
= & \frac{\mathbb{P}_{\vartheta, \eta} \left( P_i \leq q_{k+1} \cap D_k^{(i)} \right)}{q_{k+1}}
\end{align*}
\]

and, consequently, starting with \(D_1^{(i)} = C_1^{(i)}\) and proceeding step-by-step for all \(k \leq m-1\), we obtain

\[
\sum_{k=1}^{m} \frac{\mathbb{P}_{\vartheta, \eta} \left( P_i \leq q_{k+1} \cap C_k^{(i)} \right)}{q_k} \leq \frac{\mathbb{P}_{\vartheta, \eta} \left( P_i \leq q_m \cap D_m^{(i)} \right)}{q_m} = 1.
\]

Hence,

\[
\text{FDR}_{\vartheta, \eta}(\varphi^{LSU}) = \sum_{i=1}^{m} \sum_{k=1}^{q_i} \frac{1}{k} \mathbb{P}_{\vartheta, \eta} \left\{ A_k^{(i)} \right\} = \sum_{i=1}^{m} q_i \sum_{k=1}^{m} \frac{\mathbb{P}_{\vartheta, \eta} \left\{ P_i \leq q_{k+1} \cap C_k^{(i)} \right\}}{q_k} \leq \sum_{i=1}^{m} \frac{q_i}{m} = \frac{m_0}{m}.
\]

This completes the proof of the theorem.
Proof of Theorem 3.3

Straightforward calculation yields

\[ \text{FDR}_{\vartheta, \eta}(\varphi^{LSU}) = \sum_{i=1}^{m_0} \sum_{k=1}^{m} \frac{1}{k} \int \mathbb{P}_{\vartheta, \eta} \left\{ A_k^{(i)} | Z = z \right\} dF_Z(z) \]
\[ = \sum_{i=1}^{m_0} \sum_{k=1}^{m} \frac{1}{k} \int \mathbb{P}_{\vartheta, \eta} \left\{ P_i \leq q_k | Z = z \right\} \times \mathbb{P}_{\vartheta, \eta} \left\{ C_k^{(i)} | Z = z \right\} dF_Z(z) \]
\[ = \sum_{i=1}^{m_0} \sum_{k=1}^{m} \frac{1}{k} \int \mathbb{P}_{\vartheta, \eta} \left\{ P_i \leq q_k | Z = z \right\} \times \mathbb{P}_{\vartheta, \eta} \left\{ C_k^{(i)} | Z = z \right\} dF_Z(z), \]

where the random events \( A_k^{(i)} \) and \( C_k^{(i)} \) are defined in the proof of Theorem 3.1. Moreover, making use of the notation \( C_Y^{(i,z)} \), introduced in the proof of Theorem 3.1, we can express \( \text{FDR}_{\vartheta, \eta}(\varphi^{LSU}) \) by

\[ \text{FDR}_{\vartheta, \eta}(\varphi^{LSU}) = \sum_{i=1}^{m_0} \sum_{k=1}^{m} \int \mathbb{P}_{\vartheta, \eta} \left\{ P_i \leq \exp \left( -z \psi^{-1} \left( \frac{q_k}{m} \right) \right) \right\} \times \mathbb{P}_{\vartheta, \eta} \left\{ C_Y^{(i,z)} \right\} dF_Z(z) \]
\[ = \sum_{i=1}^{m_0} \sum_{k=1}^{m} \int \exp \left( -z \psi^{-1} \left( q_k \right) \right) \mathbb{P}_{\vartheta, \eta} \left\{ C_Y^{(i,z)} \right\} dF_Z(z) \]
\[ \geq \sum_{i=1}^{m_0} \sum_{k=1}^{m} \min_{k \in \{1, \ldots, m\}} \left\{ \exp \left( -z \psi^{-1} \left( q_k \right) \right) \right\} \times \mathbb{P}_{\vartheta, \eta} \left\{ C_Y^{(i,z)} \right\} dF_Z(z), \]

where the latter inequality follows from \( Y_\ell \sim \text{UNI}[0,1] \) for all \( 1 \leq \ell \leq m \) and the fact that each \( H_i \) is a true null hypothesis.
Now, it holds that

\[ F_{\text{FDR}, \varphi^{LSU}} \geq \sum_{i=1}^{m_0} \frac{q}{m} \int \min_{k \in \{1, \ldots, m\}} \left\{ \frac{\exp \left( -z \psi^{-1} \left( \frac{kq}{m} \right) \right)}{kq/m} \right\} \times \sum_{k=1}^{m} P_{\varphi, \eta} \left\{ C_{Y;k}^{(i,z)} \right\} dF_Z(z) \]

\[ = \sum_{i=1}^{m_0} \frac{q}{m} \int \min_{k \in \{1, \ldots, m\}} \left\{ \frac{\exp \left( -z \psi^{-1} \left( \frac{kq}{m} \right) \right)}{kq/m} \right\} dF_Z(z) \]

\[ = \int \min_{k \in \{1, \ldots, m\}} \left\{ \frac{\exp \left( -z \psi^{-1} \left( \frac{kq}{m} \right) \right)}{kq/m} \right\} dF_Z(z) \sum_{i=1}^{m_0} \frac{q}{m} \]

\[ = \int \min_{k \in \{1, \ldots, m\}} \left\{ \frac{\exp \left( -z \psi^{-1} \left( \frac{kq}{m} \right) \right)}{kq/m} \right\} dF_Z(z) m_0 q \frac{m}{m} . \]

This completes the proof of the theorem. \(\square\)

**Proof of Theorem 3.4**

Let

\[ \tilde{P}_i = \begin{cases} P_i, & \text{if } H_i \text{ is true}, \\ F_i(P_i), & \text{if } H_i \text{ is false}. \end{cases} \]

Because \( \tilde{P}_1, \ldots, \tilde{P}_m, \ldots \) is an exchangeable sequence of random variables with values in \([0, 1]\), it exists a random variable \( Z \) with support on \([0, 1]\) and distribution function \( F_Z \) such that the joint distribution of \( \tilde{P}_1, \ldots, \tilde{P}_m \) is given by (cf. Hewitt and Savage (1955))

\[ F_{\tilde{P}_1, \ldots, \tilde{P}_m}(p_1, \ldots, p_m) = \int F_{\tilde{P}_1|Z=z}(p_1) \times \ldots \times F_{\tilde{P}_m|Z=z}(p_m) dF_Z(z). \quad (40) \]

Moreover, for any \( F_{\tilde{P}_i} \) and \( F_Z \) it exists a random variable \( Y_i \sim \text{UNI}[0, 1] \), stochastically independent of \( Z \), such that (cf. Marshall and Olkin (1988), p. 834)

\[ p_i = F_{\tilde{P}_i}(p_i) = \int F_{Y_i}(\exp(-\psi^{-1}(p_i))) dF_Z(z), \quad (41) \]

where \( F_{Y_i}(y_i) = y_i \) and \( \psi \) is the Laplace transform of \( Z \), i. e., \( \psi(t) = \mathbb{E}[\exp(-tZ)] \). A stochastic representation of \( (\tilde{P}_i : 1 \leq i \leq m) \) is then given by

\[ (\tilde{P}_i : 1 \leq i \leq m) \overset{d}{=} (\psi(\log Y_i^{-1/Z}) : 1 \leq i \leq m). \quad (42) \]

Since \( Y_i \) and \( Z \) are independent, the conditional distribution of \( \tilde{P}_i|Z = z \) is equal to the distribution of \( \psi(\log Y_i^{-1/z}) \), which leads to

\[ F_{\tilde{P}_i|Z=z}(p_i) = F_{Y_i}(\exp(-\psi^{-1}(p_i))). \quad (43) \]

\[ \text{imsart-generic ver. 2011/11/15 file: FDR-AC.tex date: May 17, 2013} \]
Hence,
\[
F_{\tilde{P}_1, \ldots, \tilde{P}_m}(p_1, \ldots, p_m) = \int F_{\tilde{Y}_1} \left( \exp \left( -\psi^{-1}(p_1) \right) \right) \ldots \times \\
F_{\tilde{Y}_m} \left( \exp \left( -\psi^{-1}(p_m) \right) \right) dF_Z(z) \\
= \int \exp \left( -z \psi^{-1} \left( F_{\tilde{P}_1}(p_1) \right) \right) \times \ldots \times \\
\exp \left( -z \psi^{-1} \left( F_{\tilde{P}_m}(p_m) \right) \right) dF_Z(z) \\
= \int \exp \left( -z \sum_{i=1}^{m} \psi^{-1} \left( F_{\tilde{P}_i}(p_i) \right) \right) dF_Z(z) \\
= \psi \left( \sum_{i=1}^{m} \psi^{-1} \left( F_{\tilde{P}_i}(p_i) \right) \right),
\]

since the last integral is the Laplace transform of $Z$ at $\sum_{i=1}^{m} \psi^{-1} \left( F_{\tilde{P}_i}(p_i) \right)$.

Finally, noting that $P_1, \ldots, P_m$ are obtained by componentwise increasing transformations of $\tilde{P}_1, \ldots, \tilde{P}_m$ we conclude the assertion. \hfill \Box

Acknowledgments

This research is partly supported by the Deutsche Forschungsgemeinschaft via the Research Unit 1735 "Structural Inference in Statistics: Adaptation and Efficiency".

References

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. 57 289-300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 29 1165-1188.

Blanchard, G., Dickhaus, T., Roquain, E. and Villers, F. (2013). On least favorable configurations for step-up-down tests. Statistica Sinica, forthcoming.

Cai, T. T. and Jin, J. (2010). Optimal rates of convergence for estimating the null density and proportion of nonnull effects in large-scale multiple testing. Ann. Stat. 38 100-145.

Chambers, J. M., Mallows, C. L. and Stuck, B. W. (1976). A method for simulating stable random variables. J. Am. Stat. Assoc. 71 340-344.

Delattre, S. and Roquain, E. (2011). On the false discovery proportion convergence under Gaussian equi-correlation. Stat. Probab. Lett. 81 111-115.

Dickhaus, T. and Gierl, J. (2013). Simultaneous test procedures in terms of p-value copulae. Proceedings on the 2nd Annual International Conference on Computational Mathematics, Computational Geometry & Statistics (CMCGS 2013) 2 75-80. Global Science and Technology Forum (GSTF).
Dudoit, S. and van der Laan, M. J. (2008). *Multiple testing procedures with applications to genomics*. Springer Series in Statistics. New York, NY: Springer.

Fengler, M. R. and Okhrin, O. (2012). Realized Copula SFB 649 Discussion Paper report No. 2012-034, Sonderforschungsbereich 649, Humboldt-Universität zu Berlin, Germany. available at http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2012-034.pdf.

Finner, H., Dickhaus, T. and Roters, M. (2007). Dependency and false discovery rate: Asymptotics. *Ann. Stat.*, **35** 1432-1455.

Finner, H., Dickhaus, T. and Roters, M. (2009). On the false discovery rate and an asymptotically optimal rejection curve. *Ann. Stat.* **37** 596-618.

Finner, H. and Roters, M. (1998). Asymptotic comparison of step-down and step-up multiple test procedures based on exchangeable test statistics. *Ann. Stat.* **26** 505-524.

Genest, C., Nešlehová, J. and Ben Ghorbal, N. (2011). Estimators based on Kendall’s tau in multivariate copula models. *Aust. N. Z. J. Stat.* **53** 157-177.

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** 499-517.

Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control. *Ann. Stat.* **32** 1035-1061.

Guo, W. and Rao, M. B. (2008). On control of the false discovery rate under no assumption of dependency. *J. Stat. Plann. Inference* **138** 3176-3188.

Hewitt, E. and Savage, L. J. (1955). Symmetric measures on Cartesian products. *Trans. Am. Math. Soc.* **80** 470-501.

Hofert, M., Mächler, M. and McNeil, A. J. (2012). Likelihood inference for Archimedean copulas in high dimensions under known margins. *J. Multivariate Anal.* **110** 133-150.

Jin, J. and Cai, T. T. (2007). Estimating the null and the proportion of nonnull effects in large-scale multiple comparisons. *J. Am. Stat. Assoc.* **102** 495-506.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *J. Multivariate Anal.* **94** 401-419.

Kanter, M. (1975). Stable densities under change of scale and total variation inequalities. *Ann. Probab.* **3** 697-707.

Kingman, J. F. C. (1978). Uses of exchangeability. *Ann. Probab.* **6** 183-197.

Marshall, A. W. and Olkin, I. (1988). Families of multivariate distributions. *J. Am. Stat. Assoc.* **83** 834-841.

McNeil, A. J. and Nešlehová, J. (2009). Multivariate Archimedean copulas, d-monotone functions and $\ell_1$-norm symmetric distributions. *Ann. Stat.* **37** 3059-3097.

Meinshausen, N., Maathuis, M. H. and Bühlmann, P. (2011). Asymptotic optimality of the Westfall-Young permutation procedure for multiple testing under dependence. *Ann. Stat.* **39** 3369-3391.

Müller, A. and Scarsini, M. (2005). Archimedean copulæ and positive dependence. *J. Multivariate Anal.* **93** 434-445.
Nelsen, R. B. (2006). An introduction to copulas. 2nd ed. Springer Series in Statistics. New York, NY: Springer.
Pollard, K. S. and van der Laan, M. J. (2004). Choice of a null distribution in resampling-based multiple testing. J. Stat. Plann. Inference 125 85-100.
Sarkar, S. K. (2002). Some results on false discovery rate in stepwise multiple testing procedures. Ann. Stat. 30 239-257.
Shorack, G. R. and Wellner, J. A. (1986). Empirical processes with applications to statistics. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York. MR838963 (88e:60002)
Storey, J. D. (2002). A direct approach to false discovery rates. J. R. Stat. Soc., Ser. B, Stat. Methodol. 64 479-498.
Sun, W. and Cai, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. J. Am. Stat. Assoc. 102 901-912.
Westfall, P. H. and Young, S. S. (1993). Resampling-based multiple testing: examples and methods for p-value adjustment. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics. Wiley, New York.