Dickhaus, Thorsten; Stange, Jens

Working Paper

Multiple point hypothesis test problems and effective numbers of tests

SFB 649 Discussion Paper, No. 2012-041

Provided in Cooperation with:
Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

Suggested Citation: Dickhaus, Thorsten; Stange, Jens (2012) : Multiple point hypothesis test problems and effective numbers of tests, SFB 649 Discussion Paper, No. 2012-041, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:
https://hdl.handle.net/10419/79597

SFB 649 Discussion Paper 2012-041

Multiple point hypothesis test problems and effective numbers of tests

Thorsten Dickhaus*
Jens Stange*

* Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

Multiple point hypothesis test problems and effective numbers of tests*

Thorsten Dickhaus

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

and

Jens Stange

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

Abstract: We consider a special class of multiple testing problems, consisting of M simultaneous point hypothesis tests in local statistical experiments. Under certain structural assumptions the global hypothesis contains exactly one element ϑ^* (say), and ϑ^* is least favourable parameter configuration with respect to the family-wise error rate (FWER) of multiple single-step tests, meaning that the FWER of such tests becomes largest under ϑ^*.

Furthermore, it turns out that concepts of positive dependence are applicable to the involved test statistics in many practically relevant cases, in particular, for multivariate normal and chi-squared distributions. Altogether, this allows for a relaxation of the adjustment for multiplicity by making use of the intrinsic correlation structure in the data. We represent product-type bounds for the FWER in terms of a relaxed Šidák-type correction of the overall significance level and compute "effective numbers of tests".

Our methodology can be applied to a variety of simultaneous location parameter problems, as in analysis of variance models or in the context of simultaneous categorical data analysis. For example, simultaneous chi-square tests for association of categorical features are ubiquitous in genome-wide association studies. In this type of model, Moskvina and Schmidt (2008) gave a formula for an effective number of tests utilizing Pearson's haplotypic correlation coefficient as a linkage disequilibrium measure. Their result follows as a corollary from our general theory and will be generalized.

JEL classification: C12, C44.

AMS 2000 subject classifications: Primary 62J15, 62F05; secondary 62F03, 60E05.
Keywords and phrases: Monotonically sub-Markovian, multiple test, multiplicity correction, multivariate chi-squared distribution, positive orthant dependence, Šidák correction, single-step test, subset pivotality.

*This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

1. Introduction

Simultaneous statistical inference and, in particular, multiple statistical hypothesis testing has become a major branch of mathematical and applied statistics during the past 20 years, cf. [1] for some bibliometric details. This growing interest is not least due to the novel challenges posed by the need to analyze ultra high-dimensional data from genetic applications. Consider, for instance, genome-wide association studies. In these, it is common to evaluate hundreds of thousands of genetic markers simultaneously with respect to their association with a given phenotype. For the theory of multiple tests, one major resulting problem is that many classical multiple test procedures or, equivalently, the corresponding adjustments for multiplicity of the overall significance level lead to extremely small local significance levels if (strong) control of the family-wise error rate (FWER) is targeted. This implies extremely low power for detecting true effects. In [2], it was proposed to relax the type I error criterion, to allow for a few false rejections and to control the expected proportion of false significances. The mathematical formalization of this idea, the false discovery rate (FDR) has proven attractive for practitioners and the so-called "Benjamini-Hochberg correction" can meanwhile be found in many statistical software packages. However, in cases with strong dependencies among test statistics or p-values, respectively, it has been shown that, even for large systems of hypotheses, the false discovery proportion (FDP) is typically not well concentrated around its expectation, the FDR (see, for example, [11] for the case of positively dependent, exchangeable test statistics). Consequently, FDR control in such a setting does not imply any type I error control guarantee for the actual experiment at hand, although positive dependency in the sense of multivariate total positivity of order 2 (MTP_2) or positive regression dependency on subsets (PRDS) ensures FDR-control of the linear step-up test proposed in [2], as proved independently in [3] and [21].

Such a counter-intuitive behavior cannot occur for FWER-controlling single-step tests, the main objects of the present work. As shown in [20], the MTP_2 property is also useful for control of the FWER. More specifically, the critical values introduced by Simes ([23]) can be used as the basis for an FWER-controlling closed test procedure, provided that the joint distribution of test statistics or p-values, respectively, is MTP_2. This result allows to improve a closed Bonferroni test uniformly. However, the result is generic in the sense that, apart from MTP_2, no further (concrete) properties of the multivariate joint distribution of test statistics are exploited.

We will demonstrate how finer-grained knowledge about the second- (or higher-) order positive dependency structure of the underlying joint distribution of test statistics can be utilized to establish a non-trivial lower bound for the amount of possible multiplicity adjustment relaxation in comparison with the independent case. This will mathematically be formalized by the "effective number of tests" of degree i, $M_{\mathrm{eff.}}^{(i)}$ for short. For $i \geq 2$, the number $M_{\mathrm{eff.}}^{(i)}$ measures the "degree of positive dependency" in the sense that a relaxed Šidák-type (cf. [22]) multiplicity adjustment with M replaced by $M_{\mathrm{eff.}}^{(i)} \leq M$ controls the FWER, where M denotes the total number of tests to be performed. The term

"effective number of tests" is already used for a longer time in the context of genetic epidemiology and genome-wide association studies (cf. our bibliographic references in Section 4), but a formal mathematical foundation of it is yet lacking.

In this work, we will restrict attention to two-sided marginal tests, because corresponding test statistics often exhibit positive correlations, even if their counterparts for the one-sided case are negatively correlated. Especially, we will be considered with distributions that are monotonically sub-Markovian with respect to lower orthants in the sense of [4].

The paper is organized as follows. In Section 2, we set up the necessary notational framework and introduce important structural properties regarding multiple point hypothesis test problems (MPHTPs), multiple tests and multivariate distributions. Section 3 shows how these properties can be combined to calculate the effective number of tests of degree i. Applications of this general theory are provided in Section 4. We conclude with a discussion in Section 5. Our own contributions will be denoted as theorems, while we label reported results from the literature as propositions.

2. Notation and Preliminaries

Throughout the work, we let $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ denote a statistical experiment and $(S, \mathcal{S})$ a measurable space with S a subset of $\mathbb{R}$. We identify hypotheses with non-empty subsets of the parameter space Θ. The tuple $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}, \mathcal{H})$ denotes a multiple test problem, where $\mathcal{H} = (H_i, 1 \leq i \leq M)$ defines a finite family of null hypotheses. We fix the cardinality M of $\mathcal{H}$ for the remainder of the work and suppress it notationally wherever possible to increase readability. The resulting alternative hypotheses are denoted by $K_i = \Theta \setminus H_i$, $1 \leq i \leq M$. The intersection hypothesis $H_0 = \bigcap_{i=1}^M H_i$ will occasionally be referred to as global hypothesis. For a given $\vartheta \in \Theta$, we will denote the index set of true null hypotheses in $\mathcal{H}$ by $I_0 \equiv I_0(\vartheta) = \{1 \leq i \leq M : \vartheta \in H_i\}$. A (non-randomized) multiple test is a measurable mapping $\varphi = (\varphi_i)_{1\leq i\leq M} : \mathcal{X} \to \{0,1\}^M$ the components of which have the usual interpretation of a statistical test for H_i versus K_i. The family-wise error rate, FWER for short, of a multiple test φ is (for a given $\vartheta \in \Theta$) defined as the probability under ϑ of at least one type I error, i. e., $\mathrm{FWER}_\vartheta(\varphi) = \mathbb{P}_\vartheta\left(\bigcup_{i\in I_0(\vartheta)}\{\varphi_i = 1\}\right)$ and φ is said to control the FWER at a pre-specified level $\alpha \in (0,1)$ if $\sup_{\vartheta\in\Theta} \mathrm{FWER}_\vartheta(\varphi) \leq \alpha$.

Under this general framework, our main objects of interest are defined as follows.

Definition 2.1 (MPHTP, VOLTS and SCRAT)**.** *Assume that $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ can be decomposed into local statistical experiments in the sense that there exist statistical experiments $(\mathcal{X}_j, \mathcal{F}_j, (\mathbb{P}_{\vartheta_j})_{\vartheta_j\in\Theta_j})_{1\leq j\leq M}$ such that*

$$\mathcal{X} = \bigtimes_{j=1}^M \mathcal{X}_j,\ \mathcal{F} = \bigotimes_{j=1}^M \mathcal{F}_j,\ \Theta = \bigtimes_{j=1}^M \Theta_j,\ \mathbb{P}_{\vartheta_j}(A_j) = \mathbb{P}_\vartheta(\pi_j^{-1}(A_j)) \text{ for } A_j \in \mathcal{F}_j,$$

where $\pi_j : (\mathcal{X}, \mathcal{F}) \to (\mathcal{X}_j, \mathcal{F}_j)$ denotes the projection on the j-th coordinate. Then we call

a) $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}, \mathcal{H})$ *a multiple point hypothesis testing problem (MPHTP), if for all* $1 \leq j \leq M$ *we have* $H_j = \{\vartheta \in \Theta \,|\, \vartheta_j = \vartheta_j^*\}$ *for pre-defined values* ϑ_j^**, and* H_0 *contains exactly one element* $\vartheta^* = (\vartheta_1^*, \ldots, \vartheta_M^*) \in \Theta$*.*

b) $T = (T_1, \ldots, T_M)^\top$ *a vector of local test statistics (VOLTS), if* $\forall j \in \{1, \ldots, M\}$*,* $T_j : (\mathcal{X}_j, \mathcal{F}_j) \to (S, \mathcal{S})$ *is a measurable mapping.*

c) $\varphi \equiv \varphi(T) = (\varphi_1, \ldots, \varphi_M)$ *a single-step componentwise rectangular rejection area test (SCRAT) defined by* T*, if* T *is a VOLTS and* $\forall j \in \{1, \ldots, M\}$ *there exist rectangles* $\Gamma_j := \{s \in S \,|\, s > c_j\}$ *for fixed constants* $c_1, \ldots, c_M \in S$ *such that* $\varphi_j = \mathbf{1}_{\Gamma_j}(T_j) = \mathbf{1}_{(c_j,\infty)}(T_j)$*. We denote by* $O_j := \{\varphi_j = 0\} = \{T_j \leq c_j\}$ *the event that* H_j *is not rejected.*

MPHTPs appear naturally in various statistical applications. For instance, by means of appropriate (re-) parametrization, multiple comparisons with a control group (Dunnett contrasts) and all pairwise comparisons (Tukey contrasts) under the one-factorial analysis of variance model can be formalized as MPHTPs. A leading example for our further considerations is a multiple (two-sided) homogeneity test problem for many contingency tables, which is ubiquitous in genetic association studies with case-control setup (association between many genetic markers and a categorical or binary phenotype). In such a case, often a SCRAT (based on a multivariate chi-square distributed VOLTS) is performed. We will provide more details on the latter situation in Section 4.

In order to maintain a self-contained presentation, we now briefly recall some concepts of positive dependency for multivariate probability distributions.

Definition 2.2 (Concepts of positive dependence)**.** *Let $(\mathcal{X}, \mathcal{F}, \mathbb{P})$ be a probability space and let $T = (T_1, \ldots, T_M)^\top : \mathcal{X} \to S^M$ be a random vector. In all definitions below, $t = (t_1, \ldots, t_M)$ denotes an arbitrary element of S^M.*

(i) For $1 \leq j \leq M$, let $P_j \equiv P_j(t) = \mathbb{P}(\max_{1\leq h\leq j} T_h \leq t_h)$, $\gamma_{j,1} \equiv \gamma_{j,1}(t) = \mathbb{P}(T_j \leq t_j)$, and

$$\gamma_{j,i} \equiv \gamma_{j,i}(t) = \mathbb{P}(T_j \leq t_j \,|\, \max_{j-i+1\leq h\leq j-1} T_h \leq t_h), 1 < i \leq j.$$

Due to chain factorization, it holds $P_M = P_i \cdot \prod_{j=i+1}^M \gamma_{j,j}$ for every fixed $1 \leq i \leq M-1$. Following [4], we call $\beta_i = P_i \cdot \prod_{j=i+1}^M \gamma_{j,i}$ the product-type probability bound of degree i. Moreover, we call T sub-Markovian of degree i (SM_i), if $\gamma_{k,k} \geq \gamma_{k,i}$ for all $i+1 \leq k \leq M$, entailing that $P_M \geq \beta_i$. We call T monotonically sub-Markovian of degree i (MSM_i), if $\gamma_{k,k} \geq \gamma_{k,i} \geq \gamma_{k,i-1} \geq \ldots \geq \gamma_{k,1}$ for $k \geq i$ and $\gamma_{k,k} \geq \gamma_{k,k-1} \geq \ldots \geq \gamma_{k,1}$ for $i > k \geq 1$, entailing $P_M \geq \beta_i \geq \beta_{i-1} \geq \ldots \geq \beta_1$.

(ii) T is called positive lower orthant dependent (PLOD), if

$$\mathbb{P}(T_1 \leq t_1, \ldots, T_M \leq t_M) \geq \prod_{j=1}^M \mathbb{P}(T_j \leq t_j).$$

In other words, PLOD is equivalent to $P_M \geq \beta_1$.

(iii) T *is called multivariate totally positive of order* 2 *(*MTP_2*), if its distribution* $\mathbb{P}^T$ *on* $(S^M, \mathcal{S}^{\otimes M})$ *has a probability density function (pdf)* $f : S^M \to [0, \infty)$ *with respect to a measure* $\sigma^{\otimes M}$*, such that for all* $u, v \in S^M$*:*

$$f(u) \cdot f(v) \leq f(\min(u, v)) \cdot f(\max(u, v)),$$

where the minimum or maximum, respectively, is being taken componentwise.

(iv) T *is called positive regression dependent on a subset* I_0 *of the set of indices* $\{1, \ldots, M\}$ *(PRDS on* I_0*), if for every increasing set* $D \subset S^M$ *and for every index* $i \in I_0$

$$\mathbb{P}(T \in D \,|\, T_i = u) \quad \text{is non-decreasing in } u.$$

Therein, the set D *is called increasing if* $u_1 \in D$ *and* $u_2 \geq u_1$ *(jointly) imply* $u_2 \in D$.

As outlined in the introduction, the MTP_2 and PRDS properties have been investigated deeply in connection with test statistics for multiple test problems in previous literature. For instance, it is well-known that MTP_2 implies PRDS on any subset and that PRDS in turn implies FWER control or FDR control, respectively, of classical multiple test procedures. In Section 3, we will make use of the MSM_i property and demonstrate its relevance for computing the effective number of tests.

Checking PRDS and MSM_i in practice can be very cumbersome, because the conditional probabilities occurring in these definitions are often hard to handle. The MTP_2 property is often more convenient to deal with. The following proposition shows that there exists a hierarchy in the concepts of positive dependence introduced in parts (i) - (iii) of Definition 2.2.

Proposition 2.1. *Under the assumptions of Definition 2.2, it holds*

(i) MTP_2 *implies* MSM_{M-1}.

(ii) MSM_i *implies* MSM_h *for all* $1 \leq h \leq i$*. In particular,* MSM_i *for* $i \geq 2$ *implies PLOD.*

Proof. The assertions under (ii) are obvious and the assertion under (i) has been proven in [13]. □

Finally, we recall the following additional condition regarding the (joint) distribution of the vector of test statistics which has been introduced and extensively been made use of in [25] for resampling.

Definition 2.3 (Subset pivotality condition)**.** *Under our general framework, assume that the global hypothesis* H_0 *is non-empty. Let* $T = (T_1, \ldots, T_M)^\top : (\mathcal{X}, \mathcal{F}) \longrightarrow (S^M, \mathcal{S}^{\otimes M})$ *be a vector of test statistics (not necessarily a VOLTS). Then,* T *is said to satisfy the subset pivotality condition (SPC), if*

$$\forall \vartheta \in \Theta : \exists\, \vartheta^* \in H_0 : \; \mathbb{P}_{\vartheta}^{T_{I_0(\vartheta)}} = \mathbb{P}_{\vartheta^*}^{T_{I_0(\vartheta)}},$$

where the subvector $T_{I_0(\vartheta)}$ corresponds to the indices of true hypotheses in $\mathcal{H}$ under $\vartheta \in \Theta$.

In an informal manner, we can now summarize our main result which we will prove in the next section:

$$\text{MPHTP} + \text{SCRAT} + \text{SPC} + \text{MSM}_i \Rightarrow M_{\text{eff.}}^{(i)}. \tag{1}$$

Equation (1) means that for an MPHTP an effective number of tests $M_{\text{eff.}}^{(i)} \leq M$ in the sense mentioned in the introduction can be computed if the structural assumptions SCRAT, SPC and MSM_i can be established. Section 4 will exemplify practically relevant situations where this is the case.

3. General Theory

As in ordinary, one-dimensional test problems with composite null hypotheses, for strong control of the FWER it is helpful to determine the "worst case situation", i. e., the parameter value(s) for which the FWER becomes largest. Any such parameter value is called a least favorable parameter configuration (LFC). Our first result is concerned with finding the LFC in our setup. Lemma 3.1 states that for an MPHTP with the structural properties given in (1) weak control (under H_0) of the FWER is equivalent to strong control (under any arbitrary $\vartheta \in \Theta$) of the FWER. In other words, the LFC is located in H_0.

Lemma 3.1. *Let $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}, \mathcal{H})$ be an MPHTP with intersection hypothesis $H_0 = \{\vartheta^*\}$. Let $T : (\mathcal{X}, \mathcal{F}) \longrightarrow (S^M, \mathcal{S}^{\otimes M})$ be a VOLTS which satisfies the SPC.*

Then ϑ^* is the unique LFC for any SCRAT $\varphi = (\varphi_1, \ldots, \varphi_M)$ defined by T with respect to the FWER, i. e.,

$$\forall \vartheta \in \Theta : \; FWER_\vartheta(\varphi) \leq FWER_{\vartheta^*}(\varphi).$$

Proof. Let $\vartheta \in \Theta$ be an arbitrary parameter value with resulting index set of true hypotheses $I_0 \equiv I_0(\vartheta)$. Since φ is a SCRAT, the event that none of the true hypotheses is falsely rejected only depends on $T_{I_0} \equiv T_{I_0(\vartheta)}$. Let $O_{I_0} = \bigcap_{i\in I_0(\vartheta)} O_i$, with O_i as in Definition 2.1.(c). Utilizing the SPC, we obtain

$$\mathbb{P}_\vartheta(O_{I_0}) = \mathbb{P}_\vartheta^{T_{I_0}}\left(\bigtimes_{i\in I_0}(S\setminus\Gamma_i)\right) = \mathbb{P}_{\vartheta^*}^{T_{I_0}}\left(\bigtimes_{i\in I_0}(S\setminus\Gamma_i)\right) = \mathbb{P}_{\vartheta^*}(O_{I_0}),$$

and, consequently,

$$\text{FWER}_\vartheta(\varphi) = 1 - \mathbb{P}_\vartheta(O_{I_0}) = 1 - \mathbb{P}_{\vartheta^*}(O_{I_0}) = \mathbb{P}_{\vartheta^*}\left(\bigcup_{i\in I_0(\vartheta)}\{\varphi_i = 1\}\right).$$

From the fact that $I_0(\vartheta) \subseteq \{1, \ldots, M\}$, we conclude

$$\text{FWER}_\vartheta(\varphi) \leq \mathbb{P}_{\vartheta^*}\left(\bigcup_{j=1}^{M}\{\varphi_j = 1\}\right) = \text{FWER}_{\vartheta^*}(\varphi).$$

□

Our main theorem connects all assumptions and provides a formula for computing effective numbers of tests.

Theorem 3.1 (Effective numbers of tests)**.** *Under the assumptions of Lemma 3.1, let T fulfill the MSM_i property for some $i \geq 1$ under ϑ^*. Define cut-offs $c = (c_1, \ldots, c_M) \in S^M$ such that $\forall j \in \{1, \ldots, M\} : \mathbb{P}_{\vartheta_j^*}(\varphi = 1) = \mathbb{P}_{\vartheta_j^*}(T_j > c_j) = \alpha_{loc.}$ for a fixed local significance level $\alpha_{loc.} \in (0,1)$ in each marginal.*

(i) In case of $i \leq 2$, set $\xi(i) = 0$. Otherwise, let

$$\xi(i) = \sum_{\ell=2}^{i-1} \frac{\log(\gamma_{\ell,\ell}(c))}{\log(1-\alpha_{loc.})}.$$

Moreover, for every $i \leq j \leq M$, define

$$\kappa_j^{(i)} \equiv \kappa_j^{(i)}(\alpha_{loc.}; T_1, \ldots, T_j) = \frac{\log(\gamma_{j,i}(c))}{\log(1-\alpha_{loc.})}. \tag{2}$$

Then it holds

$$\sup_{\vartheta \in \Theta} FWER_\vartheta(\varphi) \leq 1 - (1-\alpha_{loc.})^{M_{eff.}^{(i)}} \tag{3}$$

for an "effective number of tests" of degree i, given by

$$M_{eff.}^{(i)} \equiv M_{eff.}^{(i)}(\alpha_{loc.}, T) = 1 + \xi(i) + \sum_{j=i\vee 2}^{M} \kappa_j^{(i)}. \tag{4}$$

(ii) Optimized bounds $\bar{\kappa}_j^{(i)}$ and $\bar{M}_{eff.}^{(i)}$:
If, for every permutation $\pi \in \mathcal{S}_M$, the MSM_i property is preserved if $T = (T_1, \ldots, T_M)$ is replaced by $(T_{\pi(1)}, \ldots, T_{\pi(M)})$, it is possible to optimize $\kappa_j^{(i)}$ and, consequently, $M_{eff.}^{(i)}$ in that the maximum strength of positive dependence between T_j and the preceding T_h, $1 \leq h \leq j-1$, is used. For $i = 2$, this leads to an optimized version

$$\bar{\kappa}_j^{(2)} \equiv \bar{\kappa}_j^{(2)}(\alpha_{loc.}; T_1, \ldots, T_j) = \frac{\log(\max_{k<j} \mathbb{P}_{\vartheta^*}(T_j \leq c_j \,|\, T_k \leq c_k))}{\log(1-\alpha_{loc.})}. \tag{5}$$

An optimized effective number of tests of degree i is given by $\bar{M}_{eff.}^{(i)} = 1 + \xi(i) + \sum_{j=i\vee 2}^{M} \bar{\kappa}_j^{(i)}$.

Proof. First, we apply Lemma 3.1 which yields

$$\begin{aligned} \sup_{\vartheta \in \Theta} \mathrm{FWER}_\vartheta(\varphi) &= \mathbb{P}_{\vartheta^*}\left(\bigcup_{j=1}^{M} \{\varphi_j = 1\}\right) = \mathbb{P}_{\vartheta^*}\left(\bigcup_{j=1}^{M} \{T_j > c_j\}\right) \\ &= 1 - \mathbb{P}_{\vartheta^*}\left(\bigcap_{j=1}^{M} O_j\right) = 1 - \mathbb{P}_{\vartheta^*}(O_1) \times \prod_{j=2}^{M} \gamma_{j,j}(c), \end{aligned}$$

where the probability measure $\mathbb{P}_{\vartheta^*}$ is used in all $\gamma_{j,j}(c)$.

Next, notice that $\mathbb{P}_{\vartheta^*}(O_j) = 1-\alpha_{\text{loc.}}$ for all $1 \leq j \leq M$. Application of MSM_i under ϑ^* entails (defining the value of an empty product as 1 and the value of an empty sum as 0) that

$$\begin{aligned}
\text{FWER}_{\vartheta^*}(\varphi) &\leq 1-(1-\alpha_{\text{loc.}})\prod_{\ell=2}^{i-1}\gamma_{\ell,\ell}(c)\prod_{j=i\vee 2}^{M}\gamma_{j,i}(c) \\
&= 1-\exp\left(\log(1-\alpha_{\text{loc.}})+\sum_{\ell=2}^{i-1}\log(\gamma_{\ell,\ell}(c))+\sum_{j=i\vee 2}^{M}\log(\gamma_{j,i}(c))\right) \\
&= 1-\exp\left(\log(1-\alpha_{\text{loc.}})\left[1+\xi(i)+\sum_{j=i\vee 2}^{M}\kappa_j^{(i)}\right]\right),
\end{aligned}$$

completing the proof of part (i). Part (ii) is then obvious. □

Remark 3.1.

(i) The numbers $(\kappa_j^{(i)})_{j=2}^M$ quantify the "degree of positive dependency" between the components of T. In particular, if T consists of jointly independent local test statistics, all $\kappa_j^{(i)}$ are equal to one (all M marginal tests "fully count") and we have $M_{\text{eff.}}^{(i)} = M$. The same holds true if $i = 1$. This special case has already been considered in [22]. On the other hand, if all T_j, $1 \leq j \leq M$, are perfectly correlated in the sense that for all $2 \leq j \leq M$: $\gamma_{j,i}(c) = 1$ leading to $\kappa_j^{(i)} = 0$, "effectively" only one single test is performed and we have $M_{\text{eff.}}^{(i)} = 1$. Obviously, in the general case we have $1 \leq M_{\text{eff.}}^{(i)} \leq M$.

(ii) For FWER control at a pre-specified (overall) significance level α in practical applications (assuming, for the moment, fixed given values of the $\kappa_j^{(i)}$), Theorem 3.1 suggests the following algorithm: (i) Start with a reasonable upper bound for $\alpha_{\text{loc.}}$, (ii) Iteratively, compute $M_{\text{eff.}}^{(i)}$ and decrease the value for $\alpha_{\text{loc.}}$, until the bound in (3) equals α.

(iii) Since $P_2 = \beta_2$ in the case of $M = 2$, inequality (3) is an equality for $M = i = 2$, even without any assumptions on the dependency structure between the components of T. We will use this fact in Example 3.1 below.

Example 3.1. *In order to illustrate the importance of the two-sidedness of the marginal test problems for our theory, let us consider the very simple example of two simultaneous Z-tests. More specifically, consider $M = 2$ and two normally distributed random variables Z_1 and Z_2, where $Z = (Z_1, Z_2)^\top \sim \mathcal{N}_2(\mu, \Sigma)$, with unknown mean vector μ, but known covariance matrix $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, $|\rho| < 1$. Moreover, let $\mu^* = (\mu_1^*, \mu_2^*) \in \mathbb{R}^2$. In the two-sided case, we let $\mathcal{H} = (H_1, H_2)$ with point hypotheses $H_1 = \{(\mu_1, \mu_2) \in \mathbb{R}^2 | \, \mu_1 = \mu_1^*\}$, $H_2 = \{(\mu_1, \mu_2) \in \mathbb{R}^2 | \, \mu_2 = \mu_2^*\}$. In this case, a suitable SCRAT at local significance*

level $\alpha_{loc.}$ *is given by* $\varphi = (\varphi_1, \varphi_2)$, *where* $\varphi_j = \mathbf{1}_{(c,\infty)}(|Z_j - \mu_j^*|)$, $j = 1, 2$, *with* $c = \Phi^{-1}(1 - \alpha_{loc.}/2)$ *and* Φ *denoting the cumulative distribution function (cdf) of the standard normal distribution. Letting* $T_j = |Z_j - \mu_j^*|$, *it is easy to show that*

$$\mathbb{P}_{\mu^*}(T_2 \leq c \,|\, T_1 \leq c) = 1 - \frac{2}{1 - \alpha_{loc.}} \int_{-c}^{c} \phi(x) \Phi\left(\frac{\rho x - c}{\sqrt{1 - \rho^2}}\right) dx,$$

with ϕ *the pdf of* $\mathcal{N}(0, 1)$. *Here, for given* $\alpha_{loc.}$, $\kappa \equiv \kappa_2^{(2)}$ *is a function of* $\rho \in (-1, 1)$ *solely, given by*

$$\kappa(\rho) = \frac{1}{\log(1 - \alpha_{loc.})} \log\left(1 - \frac{2}{1 - \alpha_{loc.}} \int_{-c}^{c} \phi(x) \Phi\left(\frac{\rho x - c}{\sqrt{1 - \rho^2}}\right) dx\right). \tag{6}$$

Simple calculus yields that κ *is decreasing in* $|\rho|$, *with maximum* $\kappa(0) = 1$ *and infimum* $\lim_{|\rho| \to 1} \kappa(\rho) = 0$. *Consequently, for the multiple test* φ *we get*

$$\alpha_{loc.} \leq FWER_{\mu^*}(\varphi) = 1 - (1 - \alpha_{loc.})^{1+\kappa(\rho)} \leq 1 - (1 - \alpha_{loc.})^2,$$

with equalities if and only if $Z_1 \overset{\mathcal{D}}{=} Z_2$, *or* Z_1 *and* Z_2 *are stochastically independent, respectively. Plainly phrased, this means that any non-zero correlation* ρ *allows an improvement of the Šidák-corrected local significance level in this two-sided case. This calculation may be regarded as a proof for the fact that the absolute values of any bivariate normal distribution are positive quadrant dependent (i. e., PLOD) for the special case that both quadrant bounds are identical.*

Contrarily, consider the two one-sided hypotheses $H_j : \{\mu_j \leq \mu_j^*\}$ *and, consequently,* $K_j : \{\mu_j > \mu_j^*\}$, $j = 1, 2$. *Now, one would choose as a suitable SCRAT at local significance level* $\alpha_{loc.}$ *the test* $\varphi = (\varphi_1, \varphi_2)$, *where* $\varphi_j = \mathbf{1}_{(c,\infty)}(Z_j - \mu_j^*)$, $j = 1, 2$, *with* $c = \Phi^{-1}(1 - \alpha_{loc.})$. *Clearly, our theory does not apply here, and negative values of* ρ *even require a stronger multiplicity correction for control of the FWER than in the Šidák case, where we have* $\alpha_{loc.} = 1 - (1 - \alpha)^{1/2}$. *To see this, consider the case* $\mu = \mu^*$ *in the global hypothesis (which is not simple in the one-sided case). We obtain, with* $T_j = Z_j - \mu_j^*$, $j = 1, 2$, *that*

$$FWER_{\mu^*}(\varphi) = \mathbb{P}_{\mu^*}(\{T_1 > c\} \cup \{T_2 > c\}) = 1 - F_{\mathcal{N}_2(0,\Sigma)}(c, c) = h(\rho) \quad \textit{(say)}.$$

Here, we get $\lim_{\rho \to 1} h(\rho) = \alpha_{loc.}$, *while we have* $\lim_{\rho \to -1} h(\rho) = 2\alpha_{loc.}$, *hence, in the worst case, even a Bonferroni correction (setting* $\alpha_{loc.} = \alpha/2$*) is necessary.*

Of course, Example 3.1 is not convincing for applying our theory, because the full joint distribution of the test statistics under H_0 is available in explicit form and, consequently, the exact calculation of the value of $\alpha_{\text{loc.}}$ required for FWER control would in practice not take the detours via our Šidák-type bound making use of $M_{\text{eff.}}^{(2)}$. However, some of the calculations in Example 3.1 will prove useful for treating more complicated situations which we will investigate in the next section.

4. Applications

4.1. Genetic association studies

In genetic association studies, a (often large) number M of positions on the human genome is simultaneously tested with respect to their association with a binary outcome (typically, a disease indicator). One study design is the case-control design, which retrospectively samples a fixed number $n_{1.}$ of cases (diseased individuals) and $n_{2.}$ healthy controls. We denote the total sample size by $n = n_{1.} + n_{2.}$. Technically, the genetic positions $1 \leq j \leq M$ are marked by single nucleotide polymorphisms (SNPs). A SNP can be described by a pair of two alleles $A_1, A_2 \in \{A, C, G, T\}$ (the single bases on the two strands of deoxyribonucleic acid at this position). Here, we restrict attention to bi-allelic SNPs, meaning that exactly two of the four letters A, C, G, T are possible at every position under investigation.

We start with a rather simple disease model, assuming that presence of one allele (the risk allele) alone may already confer an increased disease risk. This leads to high-dimensional categorical data analysis in (2×2) contingency tables. In Definition 4.1, we make the convention that the table rows correspond to the binary disease status (diseased / healthy) and the columns to the allele types (risk allele / wild type allele). In order to keep notation feasible, we abuse it a little bit in Definition 4.1 and denote the number of observational units (alleles) in the study, in the subsample of cases and in the subsample of controls, respectively, by n, $n_{1.}$ and $n_{2.}$.

Definition 4.1 (Multiple allelic association test problem)**.** *Consider an MPHTP with local statistical experiments* $(\mathcal{X}_j, \mathcal{F}_j, (\mathbb{P}_{\vartheta_j})_{\vartheta_j \in \Theta_j})_{1 \leq j \leq M}$*, such that* $\mathcal{X}_j = \mathbb{N}^{2\times 2}$ *and* $\mathcal{F}_j = 2^{\mathcal{X}_j}$*. An observation* $\mathbf{x}_j = \begin{pmatrix} x_{11}^{(j)} & x_{12}^{(j)} \\ x_{21}^{(j)} & x_{22}^{(j)} \end{pmatrix} \in \mathcal{X}_j$ *necessarily fulfills* $x_{11}^{(j)} + x_{12}^{(j)} = n_{1.}$ *and* $x_{21}^{(j)} + x_{22}^{(j)} = n_{2.}$ *by experimental design. Denoting the multinomial distribution with c categories, sample size n and vector of probabilities p by* $\mathcal{M}_c(n, p)$*, we have that, for every j, the pair of random variables* $(X_{11}^{(j)}, X_{12}^{(j)})$ *is distributed as* $\mathcal{M}_2(n_{1.}, p_j)$*, with* $p_j = (p_{1j}, p_{2j})$ *taking the role of* ϑ_j *in our general setup. The point hypothesis* H_j *that we are concerned with is then given by* $p_j = P_j$*, where* $P_j = (P_{1j}, P_{2j})$ *denotes the vector of (expected) allele frequencies at position j in the entire population (which is unknown in practice). We refer to this MPHTP as a multiple allelic association test problem. Letting* $n_{.1}^{(j)} = x_{11}^{(j)} + x_{21}^{(j)}$*,* $n_{.2}^{(j)} = x_{12}^{(j)} + x_{22}^{(j)}$*,* $\hat{p}_{1j} = n_{.1}^{(j)}/n$*,* $\hat{p}_{2j} = n_{.2}^{(j)}/n$*, a VOLTS for this MPHTP is given by* $T = (T_1, \ldots, T_M)$ *with*

$$T_j = \frac{(X_{11}^{(j)} - n_{1.}\hat{p}_{1j})^2}{n_{1.}\hat{p}_{1j}} + \frac{(X_{12}^{(j)} - n_{1.}\hat{p}_{2j})^2}{n_{1.}\hat{p}_{2j}} = \frac{(X_{11}^{(j)} - n_{1.}\hat{p}_{1j})^2}{n_{1.}\hat{p}_{1j}\hat{p}_{2j}}, 1 \leq j \leq M. \quad (7)$$

Under H_j*,* T_j *is asymptotically* $(n \to \infty)$ χ_1^2*-distributed, and we have a SCRAT at asymptotic local significance level* $\alpha_{loc.}$*, given by* $\varphi = (\varphi_j : 1 \leq j \leq M)$ *with* $\varphi_j = \mathbf{1}_{(c,\infty)}(T_j)$*, where* $c = F_{\chi_1^2}^{-1}(1 - \alpha_{loc.})$.

Such multiple allelic association test problems have already been considered in [18] with respect to the effective number of tests. As a preparation for our investigations regarding the genotypic test problem below, we briefly summarize how to compute effective numbers of tests for a multiple allelic association test problem. To this end, it is convenient to formalize the problem from a slightly different perspective: In principle, 2^M haplotypes (combinations of all M SNP-wise alleles) are possible. We assume that the (expected) haplotype frequencies in the population can be described by a probability measure π on $\{1,\ldots,2^M\}$, with $\pi_\ell \equiv \pi(\{\ell\})$ denoting the probability that a randomly chosen individual exhibits haplotype ℓ. Without loss of generality, assume that $\pi_\ell > 0$ for $1 \leq \ell \leq h$ with $\sum_{\ell=1}^{h} \pi_\ell = 1$ for some $h \leq 2^M$. Let $A(j)$ denote the set of haplotypes which imply the risk allele (corresponding to the first column in the (2×2)-table) at position j. Then, $P_{1j} = 1 - P_{2j} = \sum_{\ell \in A(j)} \pi_\ell$.

By $S_{n,\ell}$, $1 \leq \ell \leq h$, we denote the random number of individuals in the sample with haplotype ℓ and set $S_n = (S_{n,1},\ldots,S_{n,h})^\top$ with values in $\mathbb{N}^h$. Analogously, we use the notation $S_{n_{1\cdot}} = (S_{n_{1\cdot},1},\ldots,S_{n_{1\cdot},h})^\top$ if only the sub-sample of cases is considered.

Lemma 4.1. *Consider the allelic association test model as in Definition 4.1.*

(i) $S_n \sim \mathcal{M}_h(n,(\pi_1,\ldots,\pi_h))$ *and* $X_{11}^{(j)} = \sum_{\ell \in A(j)} S_{n_{1\cdot},\ell}$.

(ii) Letting

$$\xi_n = \left(\frac{S_{n,1} - n\pi_1}{\sqrt{n\pi_1}}, \ldots, \frac{S_{n,h} - n\pi_h}{\sqrt{n\pi_h}} \right),$$

it holds $\xi_n \xrightarrow{\mathcal{D}} \xi \sim \mathcal{N}_h(0,C)$, $n \to \infty$, *with* $C = E_h - \eta\eta^\top$ *and* $\eta = (\sqrt{\pi_1},\ldots,\sqrt{\pi_h})^\top$, *where* E_h *denotes the identity matrix in* $\mathbb{R}^{h\times h}$.

(iii) Under the global hypothesis H_0, *there exist linearly independent unit vectors* $(v_1,\ldots,v_M) \in \mathbb{R}^h$, *each of which perpendicular to* η, *such that* $T_j \xrightarrow{\mathcal{D}} \langle \xi, v_j \rangle^2$ *for all* $1 \leq j \leq M$.

(iv) Letting

$$\forall\, 1 \leq j \leq M : Z_j = \frac{X_{11}^{(j)} - n_{1\cdot}\hat{p}_{1j}}{\sqrt{n_{1\cdot}\hat{p}_{1j}\hat{p}_{2j}}}$$

and $Z = (Z_1,\ldots,Z_M)$, *it holds* $\mathcal{L}(Z) \xrightarrow{w} \mathcal{N}_M(0,\Sigma)$ *under* H_0, *where* $\Sigma_{ij} = \langle v_i, v_j \rangle = \rho_{ij}$ *(say), and* ρ_{ij} *is equal to Pearson's haplotypic correlation coefficient of SNPs* i *and* j *(which is tabulated for several target populations).*

Proof. Part (i) is obvious. Part (ii) is due to Section 30.1 in [8]. Part (iii) is a condensed version of Appendix A in [18] and part (iv) immediately follows from part (iii) and the construction of the vectors v_j in [18]. □

Proposition 4.1. *Let* $X = (X_1,\ldots,X_M)$ *denote a centered multivariate Gaussian random vector,* $X \sim \mathcal{N}_M(0,\Sigma)$ *with* Σ *positive definite and let* $|X| = (|X_1|,\ldots,|X_M|)$.

(i) Independently of Σ, $|X|$ *is PLOD.*

(ii) $|X|$ *is* MTP_2 *if and only if there exists a diagonal matrix* D *with diagonal elements* ± 1 *such that the off-diagonal elements of* $-D\Sigma^{-1}D$ *are all non-negative.*

(iii) For any Σ *with diagonal elements all equal to some* $\sigma^2 > 0$, $P_M \geq \beta_2 \geq \beta_1$ *for* $T_j = X_j^2$, $1 \leq j \leq M$, *if all* $t_j > 0$, $1 \leq j \leq M$, *are identical and equal to* u *(say). Furthermore, the optimized version* $\bar{M}_{eff.}^{(2)}$ *of the effective number of tests of degree* 2 *is applicable in this case.*

Proof. Part (i) is Corollary 1 in [22]. Part (ii) is Theorem 3.1 in [15]. To prove part (iii), we first notice that $\mathbb{P}(X_j^2 \leq c) = \mathbb{P}(|X_j| \leq \sqrt{c})$ for all $1 \leq j \leq M$ and all $c \geq 0$. Since all bivariate marginal distributions of a normal distribution are bivariate normal and the PLOD property for the absolute values of a Gaussian random vector is valid without any assumptions on the dimension or on Σ, we have that every pair $(|X_k|, |X_\ell|)$ is PLOD. This entails $\beta_2 \geq \beta_1$, even without the extra assumption on t. Moreover, in Appendix A in [18] it is stated that for any $2 \leq j \leq M$: $\mathbb{P}(|X_j| \leq u \,|\, \max_{1 \leq h \leq j-1} |X_h| \leq u) \geq \mathbb{P}(|X_j| \leq u \,|\, |X_k| \leq u)$ for any $k < j$. Following the reasoning of Theorem 3.1.(A) in [4], we conclude the assertion of part (iii). □

Making use of our general Theorem 3.1 with $i = 2$, we obtain the main result in [18].

Corollary 4.1 (Moskvina and Schmidt, 2008)**.** *For the multiple allelic association test problem defined in Definition 4.1, we asymptotically* $(n \to \infty)$ *get*

$$\sup_{\vartheta \in \Theta} FWER_\vartheta(\varphi) \leq 1 - (1 - \alpha_{loc.})^{\bar{M}_{eff.}^{(2)}}$$

for the effective number of simultaneous χ^2*-tests of degree* 2*, i. e.,*

$$\bar{M}_{eff.}^{(2)} \equiv \bar{M}_{eff.}^{(2)}(\alpha_{loc.}, T_1, \ldots, T_M) = 1 + \sum_{j=2}^{M} \bar{\kappa}_j^{(2)},$$

where $\bar{\kappa}_j^{(2)} = \bar{\kappa}_j^{(2)}(\Sigma)$ *is computed as* $\kappa(\rho)$ *in* (6) *with* ρ *replaced by* $\max_{k<j} \rho_{jk}$.

Proof. We notice that for all $2 \leq j \leq M$ and with p^* denoting the unique element in H_0, it holds

$$\forall k < j : \mathbb{P}_{p^*}(T_j \leq c \,|\, T_k \leq c) = \mathbb{P}_{p^*}(|Z_j| \leq \sqrt{c} \;\mid\; |Z_k| \leq \sqrt{c}).$$

Since $\sqrt{c} = \Phi^{-1}(1 - \alpha_{\text{loc.}}/2)$, we can further proceed exactly as in the first part of Example 3.1, by making use of the asymptotic result in part (iv) of Lemma 4.1 and the optimized version $\bar{M}_{\text{eff.}}^{(2)}$ which is valid due to Proposition 4.1.(iii). □

In the remainder of this section, we will demonstrate that our methodology also applies for genotypic association tests, in which the allele pairs forming the SNPs are analyzed instead of the risk alleles alone. This leads to analyzing

many (2×3) contingency tables simultaneously. Thereby, the two rows again correspond to the phenotypic (disease) status and the three columns now conventionally refer to A_1A_1 (column 1), A_1A_2 (column 2) and A_2A_2, where A_1 denotes the minor allele (less prevalent in the population) and A_2 the major allele at the respective genetic position. As we will see, this requires a non-obvious extension of the considerations for multiple allelic association test problems. Moreover, in contrast to the allelic tests considered before, in Definition 4.2 every observational unit is one individual.

Definition 4.2 (Multiple genotypic association test problem)**.** *Consider an MPHTP with local statistical experiments* $(\mathcal{X}_j, \mathcal{F}_j, (\mathbb{P}_{\vartheta_j})_{\vartheta_j \in \Theta_j})_{1 \leq j \leq M}$*, such that* $\mathcal{X}_j = \mathbb{N}^{2\times 3}$ *and* $\mathcal{F}_j = 2^{\mathcal{X}_j}$*. An observation* $\mathbf{x}_j = \begin{pmatrix} x_{11}^{(j)} & x_{12}^{(j)} & x_{13}^{(j)} \\ x_{21}^{(j)} & x_{22}^{(j)} & x_{23}^{(j)} \end{pmatrix} \in \mathcal{X}_j$ *fulfills* $x_{11}^{(j)} + x_{12}^{(j)} + x_{13}^{(j)} = n_{1\cdot}$ *and* $x_{21}^{(j)} + x_{22}^{(j)} + x_{23}^{(j)} = n_{2\cdot}$ *by experimental design. For every* j*, the triple of random variables* $(X_{11}^{(j)}, X_{12}^{(j)}, X_{13}^{(j)})$ *is distributed as* $\mathcal{M}_3(n_{1\cdot}, p_j)$*, with unknown parameter vector* $p_j = (p_{1j}, p_{2j}, p_{3j})$*. The point hypothesis* H_j *that we are concerned with is then given by* $p_j = P_j = (P_{1j}, P_{2j}, P_{3j})$*, where* P_j *denotes the vector of expected genotype frequencies at position* j *in the entire population. We refer to this MPHTP as a multiple genotypic association test problem. As in Definition 4.1, we let* $n_{\cdot 1}^{(j)} = x_{11}^{(j)} + x_{21}^{(j)}$*,* $n_{\cdot 2}^{(j)} = x_{12}^{(j)} + x_{22}^{(j)}$*,* $n_{\cdot 3}^{(j)} = x_{13}^{(j)} + x_{23}^{(j)}$*, and* $\hat{p}_{ij} = n_{\cdot i}^{(j)}/n$*,* $i = 1, 2, 3$*. A VOLTS for this MPHTP is given by* $T = (T_1, \ldots, T_M)$ *where for all* $1 \leq j \leq M$ *:*

$$\begin{aligned} T_j &= \frac{(X_{11}^{(j)} - n_{1\cdot}\hat{p}_{1j})^2}{n_{1\cdot}\hat{p}_{1j}} + \frac{(X_{12}^{(j)} - n_{1\cdot}\hat{p}_{2j})^2}{n_{1\cdot}\hat{p}_{2j}} + \frac{(X_{13}^{(j)} - n_{1\cdot}\hat{p}_{3j})^2}{n_{1\cdot}\hat{p}_{3j}} \\ &= \frac{(X_{11}^{(j)} - n_{1\cdot}\hat{p}_{1j})^2}{n_{1\cdot}\hat{p}_{1j}(1-\hat{p}_{1j})} + \frac{[\hat{p}_{2j}(X_{11}^{(j)} - n_{1\cdot}\hat{p}_{1j}) + (1-\hat{p}_{1j})(X_{12}^{(j)} - n_{1\cdot}\hat{p}_{2j})]^2}{n_{1\cdot}\hat{p}_{2j}(1-\hat{p}_{1j})(1-\hat{p}_{1j}-\hat{p}_{2j})}. \end{aligned}$$

Under H_j*,* T_j *is asymptotically* $(n \to \infty)$ χ^2_2*-distributed, and we have a SCRAT at asymptotic local significance level* $\alpha_{loc.}$*, given by* $\varphi = (\varphi_j : 1 \leq j \leq M)$ *with* $\varphi_j = \mathbf{1}_{(c,\infty)}(T_j)$*, where* $c = F^{-1}_{\chi^2_2}(1 - \alpha_{loc.})$*.*

In analogy to the allelic case, 3^M haplotypes (combinations of all M SNPs) with respect to allele pairs are possible (in principle). Again, we assume a probability measure π on $\{1, \ldots, 3^M\}$ with $\pi_\ell > 0$ for $1 \leq \ell \leq g$ and $\sum_{\ell=1}^{g} \pi_\ell = 1$ for some $g \leq 3^M$ to formalize genotypic haplotype probabilities. From this, we can deduce the genotype probabilities for SNP j as

$$P_{ij} = \sum_{\ell \in B_i(j)} \pi_\ell, i = 1, 2, 3. \tag{8}$$

In (8), $B_1(j)$, $B_2(j)$ and $B_3(j)$, respectively, denote the sets of haplotypes implying the genotype A_1A_1, A_1A_2 and A_2A_2, respectively, at position j. Furthermore, we make use of the notations $S_{n,\ell}$, $1 \leq \ell \leq g$, S_n, and $S_{n_{1\cdot}} = (S_{n_{1\cdot},1}, \ldots, S_{n_{1\cdot},g})^\top$ in corresponding manner as in the allelic situation.

Lemma 4.2. *Consider the genotypic association test model as in Definition 4.2.*

(i) $S_n \sim \mathcal{M}_g(n,(\pi_1,\ldots,\pi_g))$. *Letting*

$$\xi_n = \left(\frac{S_{n,1}-n\pi_1}{\sqrt{n\pi_1}},\ldots,\frac{S_{n,g}-n\pi_g}{\sqrt{n\pi_g}}\right),$$

it holds $\xi_n \xrightarrow{\mathcal{D}} \xi \sim \mathcal{N}_g(0,C)$, $n\to\infty$, *with* $C = E_g - \eta\eta^\top$ *and* $\eta = (\sqrt{\pi_1},\ldots,\sqrt{\pi_g})^\top$.

(ii) For all $1\le j\le M$, *it holds*

$$X_{11}^{(j)} = \sum_{\ell\in B_1(j)} S_{n_1\cdot,\ell} \quad \text{and} \quad X_{12}^{(j)} = \sum_{\ell\in B_2(j)} S_{n_1\cdot,\ell}. \tag{9}$$

Let g*-dimensional vectors* $w_1(j)$ *and* $w_2(j)$ *be defined by setting for all* $1\le i\le g$ *their* i*-th entries to*

$$\begin{aligned} w_1^{(i)}(j) &= \sqrt{\pi_i}\mathbf{1}_{B_1(j)}(i), \\ w_2^{(i)}(j) &= \sqrt{\pi_i}\left(P_{2,j}\mathbf{1}_{B_1(j)}(i) + (1-P_{1,j})\mathbf{1}_{B_2(j)}(i)\right) \end{aligned}$$

and put $v_1(j) = Cw_1(j)\{||Cw_1(j)||\}^{-1}$, $v_2(j) = Cw_2(j)\{||Cw_2(j)||\}^{-1}$. *Then, we have*

$$\langle \xi_{n_1\cdot}, v_1(j)\rangle = \frac{X_{11}^{(j)} - n_{1\cdot}P_{1j}}{\sqrt{n_{1\cdot}P_{1j}(1-P_{1j})}} = Z_{1,j} \quad \textit{(say)}, \tag{10}$$

$$\begin{aligned} \langle \xi_{n_1\cdot}, v_2(j)\rangle &= \frac{P_{2j}(X_{11}^{(j)} - n_{1\cdot}P_{1j}) + (1-P_{1j})(X_{12}^{(j)} - n_{1\cdot}P_{2j})}{\sqrt{n_{1\cdot}P_{2j}(1-P_{1j})(1-P_{1j}-P_{2j})}} \\ &= Z_{2,j} \quad \textit{(say)}. \end{aligned} \tag{11}$$

Moreover, for $n\to\infty$, *we obtain that, under* H_j, $(Z_{1,j},Z_{2,j})^\top \xrightarrow{\mathcal{D}} (Z_1,Z_2)^\top$ *with* $(Z_1,Z_2)^\top \sim \mathcal{N}_2(0,E_2)$ *and* $T_j \xrightarrow{\mathcal{D}} Z_1^2+Z_2^2$.

(iii) Under the global hypothesis H_0, *it holds for all* $1\le j,k\le M$*: For any tuple* $(\ell,m)\in\{1,2\}^2$, *the joint distribution of* $(Z_{\ell,j},Z_{m,k})^\top$ *converges weakly to a bivariate normal distribution with correlation coefficient given by*

$$\lim_{n\to\infty} Cov(Z_{\ell,j},Z_{m,k}) = r_{j,k}(\ell,m) = \langle v_\ell(j), v_m(k)\rangle. \tag{12}$$

Consequently, the vector $T=(T_1,\ldots,T_M)$ *asymptotically follows a multivariate central chi-squared distribution in the sense of Definition 3.5.7 in [24], with correlation structure given by*

$$\lim_{n\to\infty} Cov(T_j,T_k) = 2\sum_{\ell=1}^{2}\sum_{m=1}^{2} r_{j,k}^2(\ell,m). \tag{13}$$

Proof. Part (i) and representation (9) are in analogy to parts (i) and (ii) in Lemma 4.1. The validity of the representations (10) and (11) can easily be verified by making use of (8), (9) and noticing that $\langle \xi_{n1\cdot}, Cw_i(j)\rangle = \langle \xi_{n1\cdot}, w_i(j)\rangle$, $i = 1, 2$, $||Cw_1(j)||^2 = ||w_1(j)||^2 - \langle \eta, w_1(j)\rangle = P_{1j} - P_{1j}^2$ and $||Cw_2(j)||^2 = ||w_2(j)||^2 - \langle \eta, w_2(j)\rangle = P_{2j}(1 - P_{1j})(1 - P_{1j} - P_{2j})$. The remainder of part (ii) is an application of the central limit theorem. Finally, part (iii) follows by the asymptotic result in part (i), together with linearity of Gaussian distributions. □

Remark 4.1. *The correlations $r_{j,k}(\ell, m)$ in (12) only depend on the expected genotype frequencies P_{ij}, P_{ik}, $i = 1, 2, 3$ and the second-order joint probabilities of genotype pairs, given in Table 1. This entails validity of the SPC for a multiple genotypic association test problem.*

TABLE 1
First-order and second-order probabilities of genotypes and genotype pairs, respectively, for two positions j and k under the genotypic association test model as in Definition 4.2.

Genotypes	A_1A_1	A_1A_2	A_2A_2	
A_1A_1	$P_{11,jk}$	$P_{12,jk}$	$P_{13,jk}$	P_{1j}
A_1A_2	$P_{21,jk}$	$P_{22,jk}$	$P_{23,jk}$	P_{2j}
A_2A_2	$P_{31,jk}$	$P_{32,jk}$	$P_{33,jk}$	P_{3j}
	P_{1k}	P_{2k}	P_{3k}	

Proposition 4.2. *Let $T = (T_1, \ldots, T_M)$ follow a multivariate central chi-squared distribution with ν degrees of freedom in every marginal and with covariance matrix Σ in the sense of Definition 3.5.7 in [24], where all diagonal elements of Σ are equal to 1.*

(i) Independently of the off-diagonal elements of Σ, T is PLOD.
(ii) Under exchangeability (entailing equi-correlation), T is MTP$_2$.

Proof. To prove part (i), we notice that the distribution of T is equal to the joint distribution of the diagonal elements $S_{1,1}, \ldots, S_{M,M}$ of a Wishart-distributed random matrix $S \sim W_M(\nu, \Sigma)$. Corollary 4.1 in [9] yields the assertion. Part (ii) is a consequence of Example 3.5. in [15]. □

Finally, let us calculate the optimized values $\bar{\kappa}_j^{(2)}$ of degree 2, defined in (5), in the context of the MPHTP given in Definition 4.2. To this end, let p^* denote the unique parameter value in H_0 and notice that for all $2 \leq j \leq M$ and for all $k < j$:

$$\mathbb{P}_{p^*}(T_j \leq c \,|\, T_k \leq c) = \frac{\mathbb{P}_{p^*}(T_j \leq c, T_k \leq c)}{1 - \alpha_{\text{loc.}}} = \frac{F_{\chi_2^2(2,\rho(T_j,T_k))}(c, c)}{1 - \alpha_{\text{loc.}}},$$

where $\chi_2^2(2, \rho)$ denotes the bivariate chi-squared distribution with two degrees of freedom in both marginals and with correlation coefficient ρ between the two marginal chi-squared variates. The cdf of $\chi_2^2(2, \rho)$ is available in closed form, see, for instance, formula (4.2) in [14]. From this, it is easy to check that $F_{\chi_2^2(2,\rho)}(c, c)$

is isotone in ρ. Therefore, the maximum in (5) for any $2 \leq j \leq M$ is attained for index $k^* = \arg\max_{k<j} \rho(T_j, T_k)$ and we obtain

$$\forall 2 \leq j \leq M : \bar{\kappa}_j^{(2)} = \frac{\log\left(F_{\chi^2_2(2,\rho(T_j,T_{k^*}))}(c,c)\right)}{\log(1-\alpha_{\text{loc.}})} - 1.$$

Remark 4.2. *Various attempts to determine the "effective dimensionality" of a multivariate chi-squared distribution have been discussed in the literature before. Already in 1954, G. E. P. Box (see [6]) determined the "effective degrees of freedom" $h = h(\Sigma_T)$ of a sum of correlated chi-squared variates $T = (T_1, \ldots, T_r)$, see Theorem 3.1 in that paper. A linearized version of h has been proposed in [7] and [19] as an effective number of tests in the context of genetic epidemiology. However, the argumentation in the two latter papers is based on a heuristic without formal mathematical arguments. Moreover, the resulting values of the effective number of tests according to the method of [7] and [19] are in practice typically much larger than $\bar{M}_{\text{eff.}}^{(2)}$ from Theorem 3.1, as demonstrated in [18]. Therefore, the method from [7] and [19] can not be recommended.*

4.2. Further possible applications

Of course, multivariate normal and chi-squared distributions do not only appear in statistical genetics, but are ubiquitous in various applications. The reasoning in Example 3.1 can be applied to all kinds of multiple (two-sided) test problems for high-dimensional location parameters under asymptotic normality. In such a case, even if the correlation structure is known, it may not be feasible to derive exact thresholds for FWER control, because the necessary high-dimensional integrals are numerically intractable. For example, the `R`-package `mvtnorm` computes multivariate t- and normal probabilities up to dimension 1000, but not for higher dimensions. Therefore, it will often be more convenient to work with the second- (or higher) order approximation given by the computation of the effective number of tests of appropriate degree. We may mention here that our methodology also applies for certain cases with unknown (marginal) variances, because MTP_2 characterizations for multivariate t-distributions are also available in the literature.

In [10], an application of the multivariate chi-squared distribution in the context of multiple likelihood ratio tests for linear hypotheses is demonstrated. If the asymptotic correlation structure of such likelihood ratio statistics can be deduced as, for example, in [16], our method can readily be applied in that context, too. Multiple tests for Gaussian variances constitute another application field in which our methods can be applied in total analogy to the considerations in Section 4.1.

The MTP_2 property is well studied for a long time now and necessary and sufficient conditions for its validity have been derived for a variety of distributional classes. Proposition 2.1 yields that for all such MTP_2 distributions the product-type probability bounds of any degree apply and the precision of approximation is in such cases mainly limited by computational restrictions.

5. Discussion and Outlook

Of course, it is mathematically unsatisfactory that $P_M \geq \beta_2 \geq \beta_1$ can only be established under specific assumptions regarding the covariance structure within the framework of a multiple genotypic association test problem as presented in Definition 4.2. In high dimensions ($M \approx 5 \times 10^5$ or even $M \approx 10^6$ is typical in genome-wide association studies), these assumptions are impossible to check in practice. However, Proposition 4.2.(i) at least guarantees that $P_M \geq \beta_1$. Moreover, the positive quadrant properties implied by Propositions 4.2.(i) yield $\beta_2 \geq \beta_1$ in this case, too. Since β_2 is explicitly available in this case via the bivariate chi-squared cdf, one may work with β_2 although there is no theoretical guarantee that the FWER is kept, because the "true" effective number of tests may lie between $M_{\text{eff.}}^{(2)}$ and M. The findings of a genome-wide association study have to be replicated independently by means of a subsequent study anyway in most cases, such that the genetics community is willing to resort on notions of effective numbers of tests that are not proved to control the FWER strictly in all cases.

Moreover, it seems that general characterizations of the validity of the MSM_i property for $i \geq 2$ in the case of absolute multivariate normal distributions are still an open problem. This has been mentioned in the discussion of [5] and a related remark can be found in [12]. In view of Proposition 2.1, it may be conjectured that subsequently sharper assumptions on the covariance structure are necessary to establish the MSM_i property for increasing $i \geq 2$. It is interesting that these problems are closely related to the Gaussian correlation conjecture, see Section 2.4 in [17] and references therein.

Future research should address the obvious question of how to modify our proposed effective numbers of tests in cases where the correlation structure itself has to be estimated from data. In such a case, the two inferential problems (estimation of the dependency structure / multiple testing) have to be solved in parallel. This is a general topic in modern multiple testing research.

Acknowledgments

We like to thank Karl Michael Schmidt for personal communication regarding Proposition 4.1.(iii).

References

[1] Benjamini, Y. (2010). Simultaneous and selective inference: Current successes and future challenges. *Biom. J.* **52**, 6 (Dec), 708–721.

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

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

[4] Block, H. W., Costigan, T., and Sampson, A. R. (1992). Product-type probability bounds of higher order. *Probab. Eng. Inf. Sci.* **6**, 3, 349–370.

[5] Block, H. W., Costigan, T. M., and Sampson, A. R. (1993). Optimal second-order product probability bounds. *J. Appl. Probab.* **30**, 3, 675–691.

[6] Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems. I. Effect of inequality of variance in the one-way classification. *Ann. Math. Stat. 25*, 290–302.

[7] Cheverud, J. M. (2001). A simple correction for multiple comparisons in interval mapping genome scans. *Heredity 87*, 52–58.

[8] Cramér, H. (1946). *Mathematical methods of statistics.* Princeton Mathematical Series 9. Princeton University Press.

[9] Das Gupta, S., Olkin, I., Savage, L., Eaton, M., Perlman, M., and Sobel, M. (1972). Inequalities on the probability content of convex regions for elliptically contoured distributions. Proc. 6th Berkeley Sympos. math. Statist. Probab., Univ. Calif. 1970, 2, 241-265 (1972).

[10] Dickhaus, T. (2012). Simultaneous Statistical Inference in Dynamic Factor Models. SFB 649 Discussion Paper 2012-033, Sonderforschungsbereich 649, Humboldt Universität zu Berlin, Germany. Available at http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2012-033.pdf.

[11] Finner, H., Dickhaus, T., and Roters, M. (2007). Dependency and false discovery rate: Asymptotics. *Ann. Stat.,* **35**, 4, 1432–1455.

[12] Glaz, J. (2000). *Probability Inequalities for Multivariate Distributions with Applications to Statistics.* Chapman and Hall / CRC Press.

[13] Glaz, J. and Johnson, B. M. (1984). Probability inequalities for multivariate distributions with dependence structures. *J. Am. Stat. Assoc. 79*, 436–440.

[14] Gunst, R. F. and Webster, J. T. (1973). Density functions of the bivariate chi-square distribution. *J. Stat. Comput. Simulation 2*, 275–288.

[15] Karlin, S. and Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multivariate Anal. 10*, 467–498.

[16] Katayama, N. (2008). *Portmanteau Likelihood Ratio Tests for Morel Selection. Discussion Paper Series 2008-1.* Faculty of Economics, Kyushu University.

[17] Li, W. and Shao, Q.-M. (2001). Gaussian processes: Inequalities, small ball probabilities and applications. Shanbhag, D. N. (ed.) et al., Stochastic processes: Theory and methods. Amsterdam: North-Holland/ Elsevier. Handb. Stat. 19, 533-597 (2001).

[18] Moskvina, V. and Schmidt, K. M. (2008). On multiple-testing correction in genome-wide association studies. *Genetic Epidemiology 32*, 567–573.

[19] Nyholt, D. R. (2004). A simple correction for multiple testing for SNPs in linkage disequilibrium with each other. *Am. J. Hum. Genet. 74*, 765–769.

[20] Sarkar, S. K. (1998). Some probability inequalities for ordered MTP_2 random variables: a proof of the Simes conjecture. *Ann. Stat.* **26**, 2, 494–504.

[21] Sarkar, S. K. (2002). Some results on false discovery rate in stepwise multiple testing procedures. *Ann. Stat.* **30**, 1, 239–257.

[22] Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *J. Am. Stat. Assoc. 62*, 626–633.

[23] Simes, R. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika 73*, 751–754.

[24] Timm, N. H. (2002). *Applied multivariate analysis.* New York, NY: Springer.

[25] Westfall, P. H. and Young, S. S. (1992). *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.

SFB 649 Discussion Paper Series 2012

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "HMM in dynamic HAC models" by Wolfgang Karl Härdle, Ostap Okhrin and Weining Wang, January 2012.
002 "Dynamic Activity Analysis Model Based Win-Win Development Forecasting Under the Environmental Regulation in China" by Shiyi Chen and Wolfgang Karl Härdle, January 2012.
003 "A Donsker Theorem for Lévy Measures" by Richard Nickl and Markus Reiß, January 2012.
004 "Computational Statistics (Journal)" by Wolfgang Karl Härdle, Yuichi Mori and Jürgen Symanzik, January 2012.
005 "Implementing quotas in university admissions: An experimental analysis" by Sebastian Braun, Nadja Dwenger, Dorothea Kübler and Alexander Westkamp, January 2012.
006 "Quantile Regression in Risk Calibration" by Shih-Kang Chao, Wolfgang Karl Härdle and Weining Wang, January 2012.
007 "Total Work and Gender: Facts and Possible Explanations" by Michael Burda, Daniel S. Hamermesh and Philippe Weil, February 2012.
008 "Does Basel II Pillar 3 Risk Exposure Data help to Identify Risky Banks?" by Ralf Sabiwalsky, February 2012.
009 "Comparability Effects of Mandatory IFRS Adoption" by Stefano Cascino and Joachim Gassen, February 2012.
010 "Fair Value Reclassifications of Financial Assets during the Financial Crisis" by Jannis Bischof, Ulf Brüggemann and Holger Daske, February 2012.
011 "Intended and unintended consequences of mandatory IFRS adoption: A review of extant evidence and suggestions for future research" by Ulf Brüggemann, Jörg-Markus Hitz and Thorsten Sellhorn, February 2012.
012 "Confidence sets in nonparametric calibration of exponential Lévy models" by Jakob Söhl, February 2012.
013 "The Polarization of Employment in German Local Labor Markets" by Charlotte Senftleben and Hanna Wielandt, February 2012.
014 "On the Dark Side of the Market: Identifying and Analyzing Hidden Order Placements" by Nikolaus Hautsch and Ruihong Huang, February 2012.
015 "Existence and Uniqueness of Perturbation Solutions to DSGE Models" by Hong Lan and Alexander Meyer-Gohde, February 2012.
016 "Nonparametric adaptive estimation of linear functionals for low frequency observed Lévy processes" by Johanna Kappus, February 2012.
017 "Option calibration of exponential Lévy models: Implementation and empirical results" by Jakob Söhl und Mathias Trabs, February 2012.
018 "Managerial Overconfidence and Corporate Risk Management" by Tim R. Adam, Chitru S. Fernando and Evgenia Golubeva, February 2012.
019 "Why Do Firms Engage in Selective Hedging?" by Tim R. Adam, Chitru S. Fernando and Jesus M. Salas, February 2012.
020 "A Slab in the Face: Building Quality and Neighborhood Effects" by Rainer Schulz and Martin Wersing, February 2012.
021 "A Strategy Perspective on the Performance Relevance of the CFO" by Andreas Venus and Andreas Engelen, February 2012.
022 "Assessing the Anchoring of Inflation Expectations" by Till Strohsal and Lars Winkelmann, February 2012.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2012

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

023 "Hidden Liquidity: Determinants and Impact" by Gökhan Cebiroglu and Ulrich Horst, March 2012.
024 "Bye Bye, G.I. - The Impact of the U.S. Military Drawdown on Local German Labor Markets" by Jan Peter aus dem Moore and Alexandra Spitz-Oener, March 2012.
025 "Is socially responsible investing just screening? Evidence from mutual funds" by Markus Hirschberger, Ralph E. Steuer, Sebastian Utz and Maximilian Wimmer, March 2012.
026 "Explaining regional unemployment differences in Germany: a spatial panel data analysis" by Franziska Lottmann, March 2012.
027 "Forecast based Pricing of Weather Derivatives" by Wolfgang Karl Härdle, Brenda López-Cabrera and Matthias Ritter, March 2012.
028 "Does umbrella branding really work? Investigating cross-category brand loyalty" by Nadja Silberhorn and Lutz Hildebrandt, April 2012.
029 "Statistical Modelling of Temperature Risk" by Zografia Anastasiadou, and Brenda López-Cabrera, April 2012.
030 "Support Vector Machines with Evolutionary Feature Selection for Default Prediction" by Wolfgang Karl Härdle, Dedy Dwi Prastyo and Christian Hafner, April 2012.
031 "Local Adaptive Multiplicative Error Models for High-Frequency Forecasts" by Wolfgang Karl Härdle, Nikolaus Hautsch and Andrija Mihoci, April 2012.
032 "Copula Dynamics in CDOs." by Barbara Choroś-Tomczyk, Wolfgang Karl Härdle and Ludger Overbeck, May 2012.
033 "Simultaneous Statistical Inference in Dynamic Factor Models" by Thorsten Dickhaus, May 2012.
034 "Realized Copula" by Matthias R. Fengler and Ostap Okhrin, Mai 2012.
035 "Correlated Trades and Herd Behavior in the Stock Market" by Simon Jurkatis, Stephanie Kremer and Dieter Nautz, May 2012
036 "Hierarchical Archimedean Copulae: The HAC Package" by Ostap Okhrin and Alexander Ristig, May 2012.
037 "Do Japanese Stock Prices Reflect Macro Fundamentals?" by Wenjuan Chen and Anton Velinov, May 2012.
038 "The Aging Investor: Insights from Neuroeconomics" by Peter N. C. Mohr and Hauke R. Heekeren, May 2012.
039 "Volatility of price indices for heterogeneous goods" by Fabian Y.R.P. Bocart and Christian M. Hafner, May 2012.
040 "Location, location, location: Extracting location value from house prices" by Jens Kolbe, Rainer Schulz, Martin Wersing and Axel Werwatz, May 2012.
041 "Multiple point hypothesis test problems and effective numbers of tests" by Thorsten Dickhaus and Jens Stange, June 2012

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".