Statistics and Probability Letters 111 (2016) 32–40

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Multiplicity- and dependency-adjusted p-values for control of the family-wise error rate

Jens Stange [a], Thorsten Dickhaus [b,*], Arcadi Navarro [c,d,e], Daniel Schunk [f]

[a] Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstraße 39, 10117 Berlin, Germany
[b] University of Bremen, Institute for Statistics, P.O. Box 330 440, 28344 Bremen, Germany
[c] Institute of Evolutionary Biology, Universitat Pompeu Fabra, Dr. Aiguader, 88, 08003 Barcelona, Spain
[d] Institució Catalana de Recerca i Estudis Avançats (ICREA), Dr. Aiguader, 88, 08003 Barcelona, Spain
[e] Center for Genomic Regulation (CRG), Dr. Aiguader, 88, 08003 Barcelona, Spain
[f] University of Mainz, Department of Economics, 55099 Mainz, Germany

ARTICLE INFO

Article history:
Received 28 June 2015
Received in revised form 4 January 2016
Accepted 4 January 2016
Available online 13 January 2016

MSC:
62J15
62P10

Keywords:
Dependency structure
Effective number of tests
Genetic epidemiology
Multiple testing
Probability approximations
Šidák correction

ABSTRACT

Under the multiple testing framework, we propose the multiplicity- and dependency-adjustment method (MADAM) which transforms test statistics into adjusted p-values for control of the family-wise error rate. For demonstration, we apply the MADAM to data from a genetic association study.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Dependency plays a crucial role in virtually all modern applications of high-dimensional data analysis, at least for two reasons. On the one hand, data generated with nowadays' high-throughput measurements typically exhibit strong temporal, spatial, or spatio-temporal dependencies due to the underlying (neuro-)biological or technological mechanisms. In biology, linkage disequilibrium for alleles and co-regulation for levels of expression of genes are two prominent examples. Hence, these dependencies have to be taken into account in any realistic statistical model for such data. On the other hand, such dependencies induce an intrinsically low-dimensional structure in the sample and/or the parameter space, thus facilitating or enabling valid statistical inference even for moderate sample sizes.

Here, we focus on the multiple testing context where $M > 1$ null hypotheses $H_1, \ldots, H_M$ are to be tested simultaneously based on one and the same data sample $x \in \mathcal{X}$. We assume that the considered multiple test procedure φ relies on test

* Corresponding author.
E-mail addresses: jens.stange.85@gmail.com (J. Stange), dickhaus@uni-bremen.de (T. Dickhaus), arcadi.navarro@upf.edu (A. Navarro), daniel.schunk@uni-mainz.de (D. Schunk).

http://dx.doi.org/10.1016/j.spl.2016.01.005
0167-7152/© 2016 Elsevier B.V. All rights reserved.

statistics $T_1, \ldots, T_M$ which are computed from x and compared with multiplicity-adjusted rejection thresholds. In prior work (see Dickhaus and Stange, 2013 and Stange et al., 2016) it has been demonstrated that classical multiple testing approaches for control of the family-wise error rate (FWER) like the Bonferroni or the Šidák correction can be improved if the distribution of the vector $\mathbf{T} = (T_1, \ldots, T_M)^\top$ exhibits strong dependencies.[1] The possible relaxation of the necessary correction for multiplicity was described by the concept of the "effective number of tests" of order i, $M_{\text{eff.}}^{(i)}$ for short; see also Section 4.3.3 of Dickhaus (2014). Roughly speaking, $M_{\text{eff.}}^{(i)}$ approximates the number of stochastically independent tests which lead to the same FWER as φ. Hence, $M_{\text{eff.}}^{(i)}$ equals M if all components $T_1, \ldots, T_M$ are stochastically independent, and it equals one if $T_1, \ldots, T_M$ are totally dependent in the sense that all of them essentially assess exactly the same information from the data sample x. Computing $M_{\text{eff.}}^{(i)}$ for $1 \leq i \leq M$ requires knowledge of the i-variate (marginal) distributions of $\mathbf{T}$ which are then utilized in (sum- or product-type) probability approximations of order i. Hence, $M_{\text{eff.}}^{(i)}$ is typically decreasing in i, because more and more information about the dependency structure is exploited.[2]

We may mention here that the term "effective number of tests" has already been used for a longer time and seems to have its origins in the field of genetic epidemiology (see the corresponding references in Dickhaus and Stange (2013)), but the foundations of this concept have to the best of our knowledge been made mathematically rigorous in Dickhaus and Stange (2013) for the first time. Methods for computing $M_{\text{eff.}}^{(3)}$ in the genetic epidemiology context have been provided in Stange et al. (2016) based on the theory of multivariate chi-square distributions; see also Dickhaus and Royen (2015).

Although $M_{\text{eff.}}^{(i)}$ describes the quantitative effect of the dependencies in the data x on the FWER behavior of φ in a transparent and straightforward manner, it has the undesirable property that it depends on the FWER level α. This is both counter-intuitive (the dependency structure is a feature only of the data sample x, not of the parameters of some method to analyze x) and inconvenient in practice, because iterative algorithms are required to match the probability approximation of order i and α for computing $M_{\text{eff.}}^{(i)}$. In the present work, we therefore introduce the multiplicity- and dependency-adjustment method of order i, MADAM$_i$ for short. The MADAM$_i$ transforms the vector $\mathbf{T}$ into a vector of p-values which are adjusted both for multiplicity and for ith order dependency. Hence, these p-values are typically larger than their unadjusted, marginal counterparts, but smaller than the Bonferroni- or Šidák-corrected marginal p-values. In addition, MADAM$_i$ does not require the specification of α, thus avoiding the undesirable properties of $M_{\text{eff.}}^{(i)}$. However, both methods are closely related by the fact that they exploit the same probability approximations of order i.

The rest of the work is structured as follows. In Section 2, the MADAM is introduced and two different variants of it are illustrated. Section 3 shows how to utilize the MADAM for evaluating genetic association studies, considering a real-data example from this field. We conclude with a discussion in Section 4. Tables displaying the numerical results for the considered real-data example are deferred to Appendix A.

2. Statistical methodology: the MADAM

2.1. Notation and preliminaries

Throughout, we assume a statistical model $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$. The null hypotheses $H_1, \ldots, H_M$ are identified with non-empty subsets of the parameter space Θ. The intersection hypothesis $H_0 = \bigcap_{j=1}^M H_j$ is called the global hypothesis. For a given $\vartheta \in \Theta$, we will denote the index set of true null hypotheses in $\mathcal{H} = \{H_1, \ldots, H_M\}$ by $I_0 \equiv I_0(\vartheta) = \{1 \leq j \leq M : \vartheta \in H_j\}$. A (non-randomized) multiple test is a measurable mapping $\varphi = (\varphi_j)_{1 \leq j \leq M} : \mathcal{X} \to \{0, 1\}^M$ the components of which have the usual interpretation of a statistical test for H_j versus K_j. The family-wise error rate of a multiple test φ is (for a given $\vartheta \in \Theta$) defined as

$$\text{FWER}_\vartheta(\varphi) = \mathbb{P}_\vartheta\left(\bigcup_{j \in I_0(\vartheta)} \{\varphi_j = 1\}\right),$$

and φ is said to (strongly) control the FWER at a pre-specified level $\alpha \in (0, 1)$ if $\sup_{\vartheta \in \Theta} \text{FWER}_\vartheta(\varphi) \leq \alpha$.

Under this general framework, we make the following assumption.

Assumption 1. There exists a parameter value $\vartheta^* \in H_0$ such that

$$\forall \vartheta \in \Theta : \ \text{FWER}_\vartheta(\varphi) \leq \text{FWER}_{\vartheta^*}(\varphi). \tag{1}$$

Thus we may assume an overall null distribution $\mathbb{P} := \mathbb{P}_{\vartheta^*}$, under which all hypotheses are true, as the worst case with respect to control of the FWER.

[1] The FWER denotes the probability for at least one type I error among the M individual tests.

[2] Mathematical conditions guaranteeing that $M_{\text{eff.}}^{(i)}$ decreases with i are provided in Dickhaus and Stange (2013).

Download English Version:

https://daneshyari.com/en/article/7549034

Download Persian Version:

https://daneshyari.com/article/7549034

Daneshyari.com