Journal of Statistical Planning and Inference 163 (2015) 21–33

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

# On stepwise control of directional errors under independence and some dependence

Wenge Guo [a,*], Joseph P. Romano [b]

[a] New Jersey Institute of Technology, Newark, NJ 07102, USA
[b] Stanford University, Stanford, CA 94305, USA

ARTICLE INFO

*Article history:*
Received 30 May 2014
Received in revised form 29 October 2014
Accepted 24 February 2015
Available online 5 March 2015

*Keywords:*
Directional error
False discovery rate
Familywise error rate
Holm procedure
Hochberg procedure
Multiple testing
Positive dependence
Stepwise procedure

ABSTRACT

In this paper, the problem of error control of stepwise multiple testing procedures is considered. For two-sided hypotheses, control of both type 1 and type 3 (or directional) errors is required, and thus mixed directional familywise error rate control and mixed directional false discovery rate control are each considered by incorporating both types of errors in the error rate. Mixed directional familywise error rate control of stepwise methods in multiple testing has proven to be a challenging problem, as demonstrated in Shaffer (1980). By an appropriate formulation of the problem, some new stepwise procedures are developed that control type 1 and directional errors under independence and various dependencies.



## 1. Introduction

The main problem considered in this paper is the construction of procedures for the simultaneous testing of $n$ parameters $\theta_i$. For convenience, the null hypotheses $\theta_i = 0$ are of interest. Of course, we would like to reject any null hypothesis if the data suitably dictates, but we also wish to make directional inferences about the signs of $\theta_i$. First, consider the problem of simultaneously testing $n$ null hypotheses against two-sided alternatives:

$$\check{H}_i : \theta_i = 0 \quad vs. \quad \check{H}_i' : \theta_i \neq 0, \quad i = 1, \ldots, n. \tag{1}$$

Suppose, for $i = 1, \ldots, n$, a test statistic $T_i$ is available for testing $\check{H}_i$. If $\check{H}_i$ is rejected, the decision regarding $\theta_i > 0$ (or $\theta_i < 0$) is made by checking if $T_i > 0$ (or $T_i < 0$). In making such rejection and directional decisions, three types of errors might occur. The first one is the usual type 1 error, which occurs when $\theta_i = 0$, but we falsely reject $\check{H}_i$ and declare $\theta_i \neq 0$. The second one is the type 2 error, which occurs when $\theta_i \neq 0$, but we fail to reject $\check{H}_i$. The last one is called type 3 or directional error, which occurs when $\theta_i > 0$ (or $\theta_i < 0$), but we falsely declare $\theta_i < 0$ (or $\theta_i > 0$). We wish to control both type 1 and type 3 errors at pre-specified levels and, subject to their control, find testing methods with small probability of type 2 errors.

Given any procedure which makes rejections as well as directional claims about any rejected hypotheses, let $\check{V}$ and $\check{S}$ denote the numbers of type 1 errors and type 3 errors, respectively, among $\check{R}$ rejected hypotheses. Let $\check{U} = \check{V} + \check{S}$ denoting

* Corresponding author.
*E-mail addresses:* wenge.guo@njit.edu (W. Guo), romano@stanford.edu (J.P. Romano).

http://dx.doi.org/10.1016/j.jspi.2015.02.009

the total number of type 1 and type 3 errors. Then, the usual familywise error rate (FWER) and false discovery rate (FDR) are defined respectively by FWER $= \Pr(\check{V} \geq 1)$ and FDR $= E\left(\check{V}/\max(\check{R}, 1)\right)$, and the mixed directional FWER and FDR are defined respectively by mdFWER $= \Pr(\check{U} \geq 1)$ and mdFDR $= E\left(\check{U}/\max(\check{R}, 1)\right)$.

The main objective of this paper is to develop stepwise procedures (described shortly) for controlling the mdFWER and mdFDR when simultaneously testing the $n$ two-sided hypotheses $\check{H}_1, \ldots, \check{H}_n$. In multiple testing, the problem of simultaneously testing $n$ two-sided hypotheses along with directional decisions subject to the control of the mdFWER is technically very challenging. Until now, only a few results have been obtained under the strong assumption of independence of the test statistics along with some additional conditions on the marginal distribution of the test statistics.

Shaffer (1980) proved that if the test statistics $T_i, i = 1, \ldots, n$ are mutually independent and if the distributions of the $T_i$'s satisfy some additional conditions, the mdFWER of a directional Holm procedure is strongly controlled at level $\alpha$. She also constructed a counterexample where the aforementioned procedure loses the control of the mdFWER even under independence when the test statistics are Cauchy distributed. Holm (1979b, 1981) extended Shaffer's (1980) result to normal distributional settings where the $T_i$'s are conditionally independent. Finner (1994) and Liu (1997) independently used Shaffer's (1980) method of proof to show the mdFWER control of directional Hochberg procedure by making the same distributional assumptions as Shaffer (1980). By generalizing Shaffer's method of proof, Finner (1999) extended Shaffer's result on the Holm procedure to a large class of stepwise or closed multiple testing procedures under the same assumptions as in Shaffer (1980). He also gave a new but very simple and elegant proof for the aforementioned result under the assumption of $TP_3$ densities. For further discussions on the mdFWER control of closed testing methods, see Westfall et al. (2013).

Another method to tackle the problem of directional errors has been considered in Bauer et al. (1986), in which the problem of testing $n$ two-sided hypotheses testing with additional directional decisions is reformulated as the problem of testing $n$ pairs of one-sided hypotheses given by

$$H_{i1} : \theta_i \leq 0 \quad \text{vs.} \quad H'_{i1} : \theta_i > 0\,,$$

and

$$\tilde{H}_{i2} : \theta_i \geq 0 \quad \text{vs.} \quad \tilde{H}'_{i2} : \theta_i < 0$$

for $i = 1, \ldots, n$. They proved that without additional distributional assumptions, only a slight improvement of the conventional Holm procedure is possible for testing these $2n$ hypotheses. They also showed by a counterexample that in general distributional settings, a further improvement of their procedure is impossible. Compared with Shaffer's (1980) directional Holm procedure for testing $n$ two-sided hypotheses, their procedure is very conservative, although it controls directional errors under more general distributional settings of arbitrary dependence.

Finally, they also reformulated the aforementioned problem as the problem of testing $n$ pairs of one-sided hypotheses given by

$$H_{i1} : \theta_i \leq 0 \quad \text{vs.} \quad H'_{i1} : \theta_i > 0,$$

and

$$H_{i2} : \theta_i > 0 \quad \text{vs.} \quad H'_{i2} : \theta_i < 0,$$

for $i = 1, \ldots, n$, among which there is exactly one true null hypothesis within each pair of one-sided hypotheses. They proved that the modified Bonferroni procedure with the critical constant $\alpha/n$ (as opposed to $\alpha/2n$) strongly controls the FWER when testing these $2n$ one-sided hypotheses. This result is of course trivial because in this formulation there are exactly $n$ true null hypotheses. At the same time, given that there are always $n$ true null hypotheses, it is perhaps surprising that one can, as we do, develop stepdown methods that improve upon this single step method. (Indeed, at any step when applying a stepdown method, there are always $n$ true null hypotheses, and this number does not reduce.)

In the above two formulations of one-sided hypotheses, there are some inherent disadvantages when developing stepwise methods for controlling the FWER. In the first formulation, there may be a different number of true null hypotheses between $\theta_i = 0$ and $\theta_i \neq 0$, which makes it challenging to develop powerful stepwise methods in this formulation, as shown in Bauer et al. (1986). In the second formulation, one possible type 1 error will not be counted even though $T_i$ is very small when $\theta_i = 0$, which makes it unable to completely control type 1 and type 3 errors in the original formulation of two-sided hypotheses even though the FWER is controlled in this formulation. Further discussion of this point will be presented later. On the other hand, the problem of the mdFDR control seems to be technically less challenging and methods for controlling the mdFDR are available (see Benjamini and Yekutieli, 2005; Guo et al., 2010).

In the next section, some basic notation is given, as well as our approach to the problem. Theorems 1–4 deal with control of the familywise error rate with directional decisions, first under independence, and then under block dependence and positive dependence. Theorems 5–8 analogously provide results for the false discovery rate.

Although many procedures are introduced in this paper, their proven control of the FWER or FDR is established under different assumptions of dependence, including independence, between-block dependence, within-block dependence, and positive dependence. It would be impossible to advocate a single procedure in applications without any knowledge of