Statistical Methodology 23 (2015) 88–102

Contents lists available at ScienceDirect

Statistical Methodology

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

Sequential tests controlling generalized familywise error rates

Shyamal K. De [a], Michael Baron [b,c,*]

[a] School of Mathematical Sciences, National Institute of Science Education and Research, Bhubaneswar, OD, India

[b] Department of Mathematics and Statistics, American University, Washington, DC, USA

[c] Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX, USA

ARTICLE INFO

Article history:
Received 29 January 2013
Received in revised form
3 October 2014
Accepted 13 October 2014
Available online 23 October 2014

MSC:
62L
62F03
62H15

Keywords:
Generalized familywise error rate
Holm–Bonferroni procedure
Likelihood ratio
Stopping boundary
Stopping time

ABSTRACT

Sequential methods are developed for conducting a large number of simultaneous tests while controlling the Type I and Type II *generalized* familywise error rates. Namely, for the chosen values of α, β, k, and m, we derive simultaneous tests of d individual hypotheses, based on sequentially collected data, that keep the probability of at least k Type I errors not exceeding level α and the probability of at least m Type II errors not greater than β. This generalization of the classical notions of familywise error rates allows substantial reduction of the expected sample size of the multiple testing procedure.

1. Introduction

A large number of sequential statistical experiments are designed to answer many questions, that is, to test a set of hypotheses. Moreover, an answer is needed for each question, and thus, each individual hypothesis has to be tested instead of one composite hypothesis. Such problems arise

* Corresponding author at: Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX, USA.
E-mail addresses: sde@niser.ac.in (S.K. De), mbaron@utdallas.edu, baron@american.edu (M. Baron).

http://dx.doi.org/10.1016/j.stamet.2014.10.001
1572-3127/

in clinical trials for testing multiple efficacy and safety endpoints [11,14,26,27], DNA and protein sequence analysis [24,30], epidemiology [10], cybersecurity [20], and so on.

For fixed-size samples, the methodology of testing multiple hypotheses is very well developed over the last two decades or so. Efficient procedures have been proposed to control the familywise error rate or the false discovery rate, see e.g., [2,8], or [7] for the overview or [3] for the bibliography.

For sequentially collected data, a few matching methods have recently been proposed for testing multiple hypotheses. An adaptive multistage step-down procedure proposed in [1] controls the *Type* I *familywise error rate*, defined as the probability of rejecting at least one true hypothesis. Generalizing the concept of Wald's sequential probability ratio test (SPRT) to multiple hypotheses, [4] develops a testing procedure that controls both Type I and Type II familywise error rates in the strong sense. By analogy, the latter is defined as the probability of accepting at least one false null hypothesis. Control of both error rates appears possible due to the flexibility of sequential designs, similarly to the single-hypothesis SPRT attaining both desired probabilities of Type I and Type II errors. A modification of this sequential procedure is proposed in [4], combining the ideas of Wald's SPRT and Holm-type stepwise testing. Improving the plain Bonferroni methods, this new algorithm requires a smaller expected sample size, reducing the overall expected costs of the experiment and at the same time controlling both familywise error rates.

Here and in the sequel, by the *sample size* we understand the *number of sampled units*, such as patients, computer parts, etc. We assume that each sampled unit i contributes to the total cost of an experiment regardless of how many components X_{ij} (such as vital signs of patients or electronic measurements of manufactured parts) are recorded on unit i. This is quite common in many experiments (e.g., [4,11,15,26]). For example, in clinical trials, certain amount is budgeted for each participating patient, covering the cost of a treatment, service, insurance, incentive, and possibly, accommodation and transportation. However, once a patient participates in the trial, the individual measurements such as items in a written or oral questionnaire, blood work, or other analysis, usually require an incomparably lower additional cost, if any at all.

Thus, we consider the cost function that is proportional to or monotonically dependent on *the number of sampling units*. It is to be distinguished from *the total number of recorded measurements* X_{ij} that is used in [1] as a *sample size* (and misinterpreted in [4,5]).

Sampling strategies should be different under these two cost functions. Under a cost function *per sampling unit*, all measurements X_{ij} will be recorded for all the sampled units. However, if a cost *per measurement* is considered, recording the jth component will probably be terminated once the answer to the corresponding jth test is obtained. The difference may be quite substantial when one of the tests requires a much larger sample size than the others.

It has been noted that the strong control of familywise error rates is an overly stringent condition in practical situations where the number of tested null hypotheses is large, such as hundreds or thousands. Examples are found in biology, genomics, computer science, communications, and many other areas, see e.g., [6,16,18,24], and many examples in [7] and [9]. Indeed, a few erroneous decisions among a large number of tests, false rejections or missed discoveries, can be tolerated. Studies show that a slight relaxation of FWER related constraints can result in a significant reduction of the required sample size.

For these reasons, [19] introduced a concept of *generalized familywise error rates* or k-FWER which is the probability of rejecting at least k true null hypotheses. Controlling k-FWER at the given desired level is a weaker constraint (for $k \geq 2$) than controlling the standard FWER, and therefore, this condition can be satisfied by a smaller sample. Several non-sequential testing procedures controlling k-FWER were developed in [12,21,22,28].

In this article, we construct sequential multiple testing procedures that control the generalized familywise error rate. Inheriting the control of both Type I and Type II errors from the original Wald's SPRT and the sequential multiple testing procedures, the proposed schemes control both *Type I k- FWER* and *Type II m-FWER*. These sequential tests are constructed by a suitable modification of sequential procedures of [5].

The concepts are formalized in the next section, and the sequential testing procedure controlling Type I and Type II FWER is reviewed. Sections 2 and 3 contain two modifications of this sequential