Journal of Statistical Planning and Inference 143 (2013) 392–399

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

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

Characterizing optimality among three-decision procedures for directional conclusions

Fredrik Jonsson

Department of Mathematics, Uppsala University, Box 480, SE-751 06 Uppsala, Sweden

ARTICLE INFO

Article history:
Received 1 December 2011
Received in revised form
14 May 2012
Accepted 3 July 2012
Available online 8 July 2012

Keywords:
Significance testing
Directional inference
Type III errors
Simultaneous inference
Optimality

ABSTRACT

Jones and Tukey (2000) proposed three-decision procedures for directional conclusions in statistical inference, considered as an alternative to the conventional usage of one- and two-tailed significance testing. Moreover, implicit in their suggestions was to consider a procedure to be optimal in case indefinite results were minimized among all procedures with a given control of error. First, we argue by example that this characterization of optimality is not very fruitful when formalized into the strong sense of uniform minimization. Next, imposing a further regularity condition on the comparative class of procedures, we relate the suggested characterization to optimality criteria from test theory (UMP unbiasedness). Similarly, we also consider characterizing optimality in terms of maximizing correct decision rates and minimizing incorrect decision rates. Finally, we demonstrate the applicability of the three considered characterizations with respect to exponential families of distributions.

1. Introduction

Jones and Tukey (2000) considered significance testing in the light of two-sample mean-comparisons (with population means μ_A and μ_B). They proposed replacing conventional two-sided t-testing of the hypothesis $H_0: \mu_A = \mu_B$ with procedures analyzing data in terms of one of three conclusions: "(*a*) *act as if* $\mu_A > \mu_B$; (*b*) *act as if* $\mu_A < \mu_B$; (*c*) *act as if the sign of* $\mu_A - \mu_B$ *is indefinite, or not (yet) determined*" (Jones and Tukey, 2000, p. 412). Moreover, they suggested the perspective that

> [. . .] a conclusion is in error only when it is "a reversal", when it asserts one direction while the (unknown) truth is the other direction (Jones and Tukey, 2000, p. 412).

Their procedure is known, in conventional terminology, as simultaneous level-α one-sided t-testing of both directions, with conclusions (a) and (b) if the corresponding opposite direction is rejected, and conclusion (c) if none of the two p-values falls below the given level α.

Note that, if the true population means are essentially equal, then the correct and the incorrect direction are both inferred with rate α with the above procedure. Hence, definite conclusions are then obtained at rate 2α. With traditional perspectives, definite conclusions, or "significant changes", reject the possibility of true changes that are vanishingly small on the given scale of measurement (cf. Shaffer, 2002, p. 356). Thus, simultaneous level-α one-sided testing is traditionally considered to be essentially equivalent to level-2α two-sided testing in this context, with "error level" 2α.

E-mail address: fredrikjonsson99@gmail.com

0378-3758/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2012.07.004

Regarding optimality, a statistical test is said to be uniformly most powerful (UMP) if non-rejection rates, at situations where the null hypothesis is false, are minimized among procedures with controlled false rejection rates. For the three-decision setting, Jones and Tukey (2000, p. 412) suggested the similar perspective:

> We want to control the rate of error, the reversal rate, while minimizing wasted opportunity, that is, while minimizing indefinite results.

In test theory, minimizing indefinite decision rates (non-rejections) on the alternative hypothesis is equivalent to maximizing correct decision rates (power), since only two possible decisions are available. However, maximizing rates of correct directional conclusion is not equivalent to minimizing rates of indefinite conclusions in the three-decision setting (cf. Section 5).

In the following we formulate a general framework for three-decision procedures (Section 2). We then consider three characterizations of optimality:

- Maximized rates of correct conclusion;
- Minimized rates of indefinite conclusion;
- Minimized rates of incorrect conclusion.

Moreover, we relate the three perspectives to well-known criteria from test theory (Sections 4–6). Finally, we consider the applicability of the three perspectives with respect to directional inferences for multi-parameter exponential families (Section 7).

1.1. *The a priori unlikelihood that the null hypothesis is true*

Jones and Tukey (2000, p. 411) commented, as a critique of two-sided *t*-testing that

> [...] point hypotheses about parameters in real-world populations that are subjected to different treatments, are always untrue when calculations are carried to enough decimal places.

A similar statement is given by Tukey (1991):

> All we know about the world teaches us that the effects of A and B are always different – in some decimal place – for any A and B (Tukey, 1991, p. 100).

Hence, it is argued in Jones and Tukey (2000) that the new proposal is more appropriate, since it "*avoids the unrealistic postulation of a null hypothesis*" (Jones and Tukey, 2000, p. 411).

It may thus seem natural, regarding mathematical formulations of the proposal, to exclude the null hypothesis from the statistical model. For instance, Shaffer (2002, p. 356) summarizes the Tukey–Jones directional proposal as follows:

> [...] the usual 2-sided, equal-tails null hypothesis test at level α can be reinterpreted as simultaneous tests of 2 directional inequality hypotheses, each at level $\alpha/2$ [...] the maximum probability of a Type I error is $\alpha/2$ if the truth of the null hypothesis is considered impossible.

Similarly, Shaffer (2004, p. 13) refers in to the proposal in terms of considering null hypotheses impossible, and consequently, in terms of excluding corresponding parameter values from statistical models.

Moreover, it seems implicit from Shaffer's (2002) comment that the application of the Tukey–Jones directional proposal requires belief in the impossibility of the null hypothesis. It was also suggested that,

> [...] not everyone may agree on the impossibility of the null hypothesis [...] (Shaffer, 2002, p. 357).

Regarding the setting of mathematical models introduced in Section 2, no precise state of affairs is excluded on the grounds of "being impossible in practice". On the other hand, where the traditional null hypothesis is considered as a specific phenomenon (of "no effect"), it now instead occurs as a boundary state of affairs with respect to two possible directions.

Confer Nickerson (2000, p. 263) for a review of related discussions (and a review of common missunderstandings) regarding the current practice of significance testing. Confer also Goudey (2007), discussing the relevance of applying the Tukey–Jones directional proposal in environmental statistics.

2. Uniformly optimal three-decision procedures

Regarding the previously mentioned example from Jones and Tukey (2000), consider the associated statistical model of normal distributions. We suggest dividing its parameter space into two regions specified by

$$H_1:\ \mu_A \leq \mu_B, \quad H_2:\ \mu_A > \mu_B. \tag{1}$$

A corresponding three-decision procedure concludes either: (a) in favor of H_1; (b) in favor of H_2; or (c) no definite decision.

Including the parameter value $\mu_A = \mu_B$ within H_1 above might be regarded as an act of convention. In fact, H_1 may be considered as a mirrored copy of H_2. Note that the distributional behaviour, regarding the elements of the considered