Statistical Methodology 21 (2014) 135–155

Contents lists available at ScienceDirect

Statistical Methodology

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

Multiple crossing sequential fixed-size confidence region methodologies for a multivariate normal mean vector

Nitis Mukhopadhyay *, Sankha Muthu Poruthotage

Department of Statistics, University of Connecticut, Storrs, CT, USA

ARTICLE INFO

Article history:
Received 21 August 2013
Received in revised form
24 December 2013
Accepted 28 March 2014

MSC:
62H12
62J15
62L12
62L05

Keywords:
Fine-tuning
First-order efficiency
Real data illustrations
Second-order efficiency
Simulation
Truncation

ABSTRACT

The *asymptotically efficient* and *asymptotically consistent* purely sequential procedure of Mukhopadhyay and Al-Mousawi (1986) is customarily used to construct a confidence region $\mathcal{R}$ for the mean vector $\boldsymbol{\mu}$ of $N_p(\boldsymbol{\mu}, \sigma^2\mathbf{H})$. This procedure does not have the *exact consistency* property. $\mathbf{H}_{p\times p}$ is assumed known and positive definite with σ^2 unknown. The maximum diameter of $\mathcal{R}$ and the confidence coefficient are prefixed.

A purely sequential sampling strategy is proposed allowing sampling until sample size crosses the boundary multiple times. We ascertain asymptotic efficiency and asymptotic consistency properties (Theorem 3.1). Its ability to nearly achieve required coverage probability without significant over-sampling is demonstrated with simulations. A truncation technique plus fine-tuning of the multiple crossing rule are proposed to increase practicality. Two real data illustrations are highlighted.

1. Introduction

In numerous applications it is crucial to compare treatments with a control. Typically, sample mean is calculated from subjects allocated to each treatment and compared with the sample mean from the

* Corresponding author. Tel.: +1 860 657 1661.
E-mail addresses: nitis.mukhopadhyay@uconn.edu, nitis_mukhopadhyay@yahoo.com (N. Mukhopadhyay), sperera@plymouthrock.com (S. Muthu Poruthotage).

http://dx.doi.org/10.1016/j.stamet.2014.03.003

control. For statistical inferences about treatment-control comparisons, a one-way ANOVA model is used:

$$Y_{ij} = \theta_i + \varepsilon_{ij} \quad \text{where } i = 1, \ldots, b,\ j = 1, \ldots, n,$$
$$\varepsilon_{ij}\text{'s are independently and identically distributed (i.i.d.) } N(0, \sigma^2), \tag{1.1}$$

where i (=1, ..., b) corresponds to the ith treatment. Each treatment is used on n subjects. Hence θ_i is the effect of the ith treatment and Y_{ij} represents the response from the jth subject on the ith treatment. Even though asserting whether the overall ANOVA is significant by testing the hypothesis of $\theta_1 = \theta_2 = \cdots = \theta_b$ may be a first step, in practical applications it is critical to determine the magnitude of the differences between the treatment effects. Recent applications come from genomic research and education.

Without any loss of generality, let $i = 1$ correspond to a control with which we want to compare all other treatments. We construct a simultaneous fixed-size ellipsoidal confidence region for $\theta_i - \theta_1$ where $i = 2, \ldots, b$ with a family-wise joint confidence coefficient $1-\alpha$, $0 < \alpha < 1$. For $i = 2, \ldots, b$, let $X_{ij} = Y_{ij} - Y_{1j}$, $\mu_i = \theta_i - \theta_1$ and $\eta_{ij} = \varepsilon_{ij} - \varepsilon_{1j}$. We rewrite (1.1):

$$X_{ij} = \mu_i + \eta_{ij} \quad \text{where } i = 2, \ldots, b,\ j = 1, \ldots, n \text{ and}$$
$$\eta_{ij}\text{'s distributed as } N(0, 2\sigma^2) \text{ random variables.} \tag{1.2}$$

Finally, let $\mathbf{X}_j = (X_{2j}, \ldots, X_{bj})'$, $\boldsymbol{\mu} = (\mu_2, \ldots, \mu_b)'$ and $\mathbf{H}$ a square matrix of dimension $(b-1)$ where all diagonal elements equal 2 and the off-diagonal elements equal 1. Then, $\mathbf{X}_j$'s, $j = 1, \ldots, n$, are i.i.d. $N_p(\boldsymbol{\mu}, \sigma^2\mathbf{H})$ with $p = b - 1$. Our problem then reduces to the construction of a fixed-size confidence region for $\boldsymbol{\mu}$ with σ^2 unknown and $\mathbf{H}$ known and positive definite (p.d.).

1.1. Fixed-width confidence interval for μ: a brief review

Stein [21,22] solved the problem of constructing a fixed-width (=2d) confidence interval with a confidence coefficient of $(1-\alpha)$ for μ, in a $N(\mu, \sigma^2)$ distribution when σ^2 is unknown, via his pioneering two-stage procedure. Subsequently, Chow and Robbins [4] investigated a purely sequential procedure which had *asymptotic consistency* and *asymptotic efficiency* properties, but it failed to have the consistency (or exact consistency) property (defined formally in (2.3)).

Mukhopadhyay [9] came up with a surprising modification of Stein's methodology that had *all* of the following first-order properties: exact consistency, asymptotic consistency, and asymptotic efficiency. This led Ghosh and Mukhopadhyay [6] to introduce the notion of *asymptotic second-order efficiency* property. For a comprehensive treatment of multi-stage procedures, one may refer to Mukhopadhyay and Solanky [17], Ghosh et al. [7], Mukhopadhyay et al. [14], and Mukhopadhyay and de Silva [15, Chapter 6], among other sources.

Mukhopadhyay and Muthu Poruthotage [16] developed the important and novel concept of multiple boundary crossing and showed its importance in the context of a fixed-width confidence interval construction. In this paper, we methodologically extend and explore that notion in the context of a fixed-size multivariate normal mean vector estimation problem.

1.2. Fixed-size confidence regions for $\boldsymbol{\mu}$: a brief review

Initially, problems of estimating the mean vector $\boldsymbol{\mu}$ of a multivariate normal density were inspired by univariate counterparts. Chatterjee [3], developed a two-stage procedure to estimate regression coefficients of a general linear model under normal errors. Mukhopadhyay and Al-Mousawi [12] introduced multi-stage and sequential fixed-size confidence region procedures and investigated their properties to estimate $\boldsymbol{\mu}$ in $N_p(\boldsymbol{\mu}, \sigma^2\mathbf{H})$.

The recent literature surrounding multivariate normal mean vector and other multi-parameter estimation problems is vast. We mention some references, for example, Aoshima et al. [1], Aoshima and Mukhopadhyay [2], and Mukhopadhyay [11] for closely related developments. Other references may be found from the cited sources.