

Accepted Manuscript

On simultaneous confidence interval estimation for the difference of paired mean vectors in high-dimensional settings

Masashi Hyodo, Hiroki Watanabe, Takashi Seo

PII: S0047-259X(17)30725-X
DOI: <https://doi.org/10.1016/j.jmva.2018.07.008>
Reference: YJMVA 4388

To appear in: *Journal of Multivariate Analysis*

Received date: 27 November 2017

Please cite this article as: M. Hyodo, H. Watanabe, T. Seo, On simultaneous confidence interval estimation for the difference of paired mean vectors in high-dimensional settings, *Journal of Multivariate Analysis* (2018), <https://doi.org/10.1016/j.jmva.2018.07.008>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



On simultaneous confidence interval estimation for the difference of paired mean vectors in high-dimensional settings

Masashi Hyodo^a, Hiroki Watanabe^b, Takashi Seo^c

^a Department of Mathematical Sciences, Graduate School of Engineering, Osaka Prefecture University,
1-1, Gakuen-cho, Naka-ku, Sakai-shi, Osaka 599-8531, Japan

^b Graduate School of Science, Tokyo University of Science 1-3 Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan

^c Department of Applied Mathematics, Tokyo University of Science,
1-3 Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan

Abstract

To test whether two populations have the same mean vector in a high-dimensional setting, Chen and Qin (2010, Ann. Statist.) derived an unbiased estimator of the squared Euclidean distance between the mean vectors and proved the asymptotic normality of this estimator under local assumptions about the mean vectors. In this study, their results are extended without assumptions about the mean vectors. In addition, asymptotic normality is established in the class of general statistics including Chen and Qin's statistics and other important statistics under general moment conditions that cover both Chen and Qin's moment condition and elliptical distributional assumption. These asymptotic results are applied to the construction of simultaneous intervals for all pair-wise differences between mean vectors of $k \geq 2$ groups. The finite-sample and dimension performance of the proposed methods is also studied via Monte Carlo simulations. The methodology is illustrated using microarray data.

AMS 2000 subject classification: Primary 62H15; secondary 62F05.

Key words: Confidence interval, High dimension, Non-normality, Statistical hypothesis testing.

1. Introduction

Testing hypotheses about mean vectors is part of many procedures in multivariate analysis. The standard approach is based on Hotelling's classical T^2 test which is known to have optimal performance properties in large samples, i.e., if we assume that the number of feature variables is fixed and is much smaller than the sample size. However, in many modern applications, the number of features exceeds the sample size, thus making the straightforward use of T^2 statistics impossible because the sample covariance matrix is singular.

Several recent papers have addressed this issue. Bai and Saranadasa [3] considered the (normal) two-sample case with equal covariance matrices, and proposed an estimator of the Euclidean norm of the shift vector instead of T^2 statistics; they also established the asymptotic normality of the test statistics assuming that the number of features and the sample size are of the same order. Nishiyama et al. [12] proposed a test procedure for linear hypotheses of a set of mean vectors from $k \geq 2$ normal populations. Chen and Qin [5] also derived a test for the two-sample problem without assuming normality. Recently, various alternative approaches for the two-sample test have been proposed; see, e.g., [4, 6, 13–15]. A good introduction to significance testing for high-dimensional mean vectors is given in [9]. In addition, k -sample significance tests for high-dimensional mean vectors have been studied in [10, 16].

Previous studies focused on testing whether the g th population mean vector μ_g and the h th population mean vector μ_h are the same or not. However, an estimate of the Euclidean norm $\|\mu_g - \mu_h\|^2$ is often needed. For example, in multivariate bioequivalence tests, the following hypotheses are of interest:

$$\mathcal{H} : \|\mu_g - \mu_h\|^2 \geq \theta \quad \text{vs.} \quad \mathcal{A} : \|\mu_g - \mu_h\|^2 < \theta.$$

Here, θ is a prescribed constant. A confidence interval for $\|\mu_g - \mu_h\|^2$ can also be used to carry out this test.

Download English Version:

<https://daneshyari.com/en/article/7546398>

Download Persian Version:

<https://daneshyari.com/article/7546398>

[Daneshyari.com](https://daneshyari.com)