Journal of Statistical Planning and Inference 175 (2016) 11–24

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Confidence regions for level differences in growth curve models

Solomon W. Harrar [a], Jin Xu [b,*]

[a] Department of Statistics, University of Kentucky, Lexington, KY 40536, USA

[b] Department of Statistics and Actuarial Science, East China Normal University, Shanghai 200241, China

ARTICLE INFO

Article history:
Received 7 February 2015
Received in revised form 31 December 2015
Accepted 18 February 2016
Available online 5 March 2016

Keywords:
Asymptotic expansion
Bartlett's correction
Characteristic function
Profile analysis
Repeated measures
Wald's criteria

ABSTRACT

In a pre-post or other kind of repeated measures study, it is sometimes clear that the mean profiles of the repeated measures are parallel across treatment groups. When for example, it can be assumed that there is no interaction between the repeated measure factor and the treatment, it would be of interest to know how much of a difference exists in the effect of the treatments. Such differences in the absence of interaction are referred to as level differences. In this paper, we consider methods for constructing confidence regions for level differences in the multi-dimensional cases. We derive asymptotic expansions for some intuitively appealing pivotal quantities to construct the confidence regions corrected up to the second order. Such corrections are shown in the multivariate literature to improve the accuracy of asymptotic approximations. We evaluate the finite sample performance of the confidence regions via a simulation study. Real-data example from forestry is used to provide an empirical illustration of the features of the various confidence regions proposed in the paper.

1. Introduction

For $j = 1, \ldots, k$, let $\mathbf{y}_i^{(j)}$, $i = 1, \ldots, n_j$, be independently distributed p-dimensional random vectors with mean $\boldsymbol{\mu}^{(j)} = (\mu_1^{(j)}, \ldots, \mu_p^{(j)})'$ and covariance matrix $\Sigma = (\sigma_{ab})$. For this paper, we assume that the components of $\mathbf{y}_i^{(j)}$ are p-repeated measurements from the same experimental/observational subject which are measured in commensurable units. In the usual repeated-measures setting, such data arises when repeated measures are taken from each of $n = n_1 + \cdots + n_k$ subjects sorted into k groups according to the population they came from or the treatment they received. In multivariate statistics, a mean profile for the jth population is a line plot connecting the points $(1, \mu_1^{(j)}), \ldots, (p, \mu_p^{(j)})$. The study of the mean profiles and their relationship with one another is known as profile analysis. In growth curve analysis, the study of the relationship between the curves $\boldsymbol{\mu}^{(j)} = X\boldsymbol{\theta}^{(j)}$ for $j = 1, \ldots, k$ is of interest where X is a $p \times q$ $(q \le p)$ known design matrix that does not vary with j and $\boldsymbol{\theta}^{(j)}$ is a q-dimensional vector of unknown parameters. Needless to say, profile analysis is a special case of growth curve analysis which can be seen by setting $q = p$ and $X = I_p$ where I_p is p-dimensional identity matrix. Multivariate approaches for growth curve analysis and, hence, profile analysis does not stipulate any structure on the covariance matrix Σ other than the necessary requirements of symmetry and positive definiteness.

The primary question of interest in growth curve analysis is whether the curves are similar or parallel. If so, the follow-up questions are whether the curves coincide and whether they are flat. An additional inferential problem of interest

* Correspondence to: Department of Statistics and Actuarial Science, East China Normal University, 500 Dongchuan Road, Shanghai 200241, China.
E-mail address: jxu@stat.ecnu.edu.cn (J. Xu).

http://dx.doi.org/10.1016/j.jspi.2016.02.005
0378-3758/

is to construct confidence interval for the distances separating the curves if similarity of the curves cannot be rejected. Mathematically, assuming

$$\boldsymbol{\mu}^{(j)} - \boldsymbol{\mu}^{(k)} = X(\boldsymbol{\theta}^{(j)} - \boldsymbol{\theta}^{(k)}) = \gamma_j \mathbf{1}_p, \quad \text{for } j = 1, \ldots, k-1, \tag{1}$$

where γ_j is some unknown real number and $\mathbf{1}_p$ is a p-dimensional vector of all 1's, interest lies in constructing confidence regions and simultaneous confidence intervals for $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{k-1})'$.

Univariate as well as multivariate methods for growth curve analysis have been around for a while. However, it was only recently that these problems were looked at from likelihood point of view. Srivastava (1987) obtained likelihood-ratio tests for the hypothesis in profile analysis. The likelihood ratio tests for coincidence and flatness hypotheses turned out to be different from what have been classically used. In addition to deriving likelihood ratio test for the three common hypotheses in profile analysis, Srivastava (1987) also considered confidence regions for $\boldsymbol{\gamma}$ by inverting the likelihood ratio tests and a T^2 simultaneous intervals (Johnson and Wichern, 2002) for $\gamma_1, \ldots, \gamma_{k-1}$. More recently, Fujikoshi (2009) and Seo et al. (2011) derived likelihood ratio tests for growth curve model. In addition to introducing an alternative method of constructing simultaneous confidence intervals, Fujikoshi (2009) also showed that the confidence regions and simultaneous intervals of Srivastava (1987) are applicable for growth curve model as well. Ohlson and Srivastava (2010) considered profile analysis of several groups where the groups have partly equal means.

It should be pointed out that both studies of Srivastava (1987) and Fujikoshi (2009) are under normality and low-dimensional situation. The focus of Srivastava (1987) is deriving likelihood ratio test and confidence regions and that of Fujikoshi (2009) is also likelihood ratio test for parallelism under general growth curve model which includes profile analysis as a special case. Fujikoshi (2009) further discusses a Wald's confidence interval for a linear combination of the level-difference parameters. In our work, we consider the high dimensional case in the sense that both sample size and dimension increase proportionally. Furthermore, in the fixed dimensional case, we introduced finitely approximated confidence region for the level difference based on asymptotic expansion of normal-theory Wald-type and likelihood ratio statistics without normality. The asymptotic expansion provides a second order approximation which is known to provide more accurate than the asymptotics.

Few recent works considered tests in profile analysis and growth curve models under non-normality (Maruyama, 2007, 2010; Harrar and Xu, 2014). Maruyama (2007) derived asymptotic expansions for the null distributions of the classical tests in profile analysis for the case of two populations ($k = 2$) which was later extended to $k \geq 2$ in Maruyama (2010). Using a different technique, Harrar and Xu (2014) derived the asymptotic expansion for the null distributions for normal-theory likelihood ratio tests as well as the classical tests for profile analysis for any $k \geq 2$. In addition, Harrar and Xu (2014) also studied tests for level and flatness hypotheses under normality as well as non-normality when the alternative hypothesis is general rather than constrained by parallelism.

The aim of the present paper is two-fold. First, we obtain high dimensional confidence region correct up to the order $p^{-3/2}$ for the case $n \geq p$ assuming both p and n diverge proportionally. For $100(1-\alpha)\%$ confidence region $C = C(\mathbf{y}_1^{(1)}, \ldots, \mathbf{y}_{n_1}^{(1)}, \ldots, \mathbf{y}_1^{(k)}, \ldots, \mathbf{y}_{n_k}^{(k)})$ for $\boldsymbol{\gamma}$, correction up to the order $p^{-3/2}$ is in the sense that $|\mathrm{Prob}(\boldsymbol{\gamma} \in C) - (1-\alpha)| = O(p^{-3/2})$. We also propose method for getting exact confidence region for level difference that can be used when $p > n$. Second, we derive confidence region for γ correct up to the order n^{-1} under non-normality.

The paper is organized as follows. In Section 2, we fix the notations and present some preliminary results used in the subsequent sections. In particular, stochastic representation for the maximum likelihood estimator of $\boldsymbol{\gamma}$ that is useful for constructing asymptotic pivotal quantity and high dimensional asymptotic expansion is given. Section 3 considers high-dimensional confidence region under normality. Section 4 deals with confidence region under non-normality with finiteness correction up to $O(n^{-1})$ based on asymptotic expansion of the null distribution of an asymptotic pivotal statistic. Numerical accuracy of the finiteness corrections are presented through simulations. Section 5 illustrates the application of the proposed methods with a real-data example. Section 6 concludes the paper with some discussions. All proofs are shifted to Appendix.

2. Preliminaries

2.1. Notations

For $j = 1, \ldots, k$, let $\bar{\mathbf{y}}^{(j)}$ and $S^{(j)}$ be the sample mean and the sample covariance matrix of the jth group, respectively. Let

$$\bar{\mathbf{Y}} = (\bar{\mathbf{y}}^{(1)}, \ldots, \bar{\mathbf{y}}^{(k)}) \quad \text{and} \quad S = \sum_{j=1}^{k} (n_j - 1) S^{(j)} / (n-k). \tag{2}$$

Define the standardized quantities

$$\mathbf{z}_j = \sqrt{n_j}\, \Sigma^{-\frac{1}{2}} (\bar{\mathbf{y}}^{(j)} - \boldsymbol{\mu}^{(j)}) \quad \text{and} \quad V_j = \sqrt{n_j} (\Sigma^{-\frac{1}{2}} S^{(j)} \Sigma^{-\frac{1}{2}} - I_p).$$

Denote $\rho_j = \sqrt{n_j/n}$, $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_k)'$,

$$Z_{p\times k} = (\mathbf{z}_1, \ldots, \mathbf{z}_k) \quad \text{and} \quad V_{p\times p} = \rho_1 V_1 + \cdots + \rho_k V_k. \tag{3}$$