Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 52 (2008) 2714–2724
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Large sample approximations for the LR statistic for equality of the smallest eigenvalues of a covariance matrix under elliptical population

Daisuke Watanabe, Susumu Okada, Yasunori Fujikoshi*, Takakazu Sugiyama

Department of Mathematics, Graduate School of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Received 9 November 2006; received in revised form 22 July 2007; accepted 26 September 2007
Available online 4 October 2007

Abstract

This paper is concerned with large sample approximations of the LR statistic for testing the hypothesis that the smallest eigenvalues of a covariance matrix are equal. Under a normal population Lawley [1956. Tests of significance for the latent roots of covariance and correlation matrices. Biometrika 43, 128–136.] and Fujikoshi [1977. An asymptotic expansion for the distributions of the latent roots of the Wishart matrix with multiple population roots. Ann. Inst. Statist. Math. 29, 379–387.] obtained a Bartlett-correction factor and an asymptotic expansion for the LR statistic, respectively, when the sample size is large. In this paper we extend the Bartlett correction factor to an elliptical population. The accuracy of our approximations is examined through simulation experiments.

Keywords: Bartlett correction; Elliptical distribution; Likelihood ratio statistic; Principal component analysis

1. Introduction

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ be a random sample of size $N = n + 1$ from a p-variate population with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let $\lambda_1 \geqslant \cdots \geqslant \lambda_p > 0$ be the eigenvalues of $\boldsymbol{\Sigma}$. We consider the problem of testing the hypothesis that the smallest $m = p - q$ eigenvalues of $\boldsymbol{\Sigma}$ are equal, i.e.,

$$\mathrm{H}_0\colon \lambda_{q+1} = \cdots = \lambda_p. \tag{1}$$

This problem is important, because in applications of principal component analysis, we are usually interested in the number of q for which the first q principal components include most of information contained within the covariance matrix. Such situation is realized if the hypothesis (1) is satisfied, the remainder are most same and small, and the qth eigenvalue and $(q + 1)$th eigenvalue are sufficiently separate.

Let S be the usual sample covariance matrix defined by

$$\mathbf{S} = \frac{1}{n}\sum_{j=1}^{N}(\boldsymbol{x}_j - \bar{\boldsymbol{x}})(\boldsymbol{x}_j - \bar{\boldsymbol{x}})', \tag{2}$$

* Corresponding author. Tel.: +81 3 3817 1745; fax: +81 3 3817 1746.
E-mail address: fujikoshi_y@yahoo.co.jp (Y. Fujikoshi).

0167-9473/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2007.09.028

where $\bar{\boldsymbol{x}} = (1/N)\sum_{j=1}^{N} \boldsymbol{x}_j$. The likelihood ratio test under normality rejects H_0 for large values of

$$T_{\mathrm{LR}} = -2\log\left(\frac{\prod_{j=q+1}^{p}\phi_i(\mathbf{S})}{\{(1/m)\sum_{j=q+1}^{p}\phi_i(\mathbf{S})\}^m}\right)^{n/2}, \tag{3}$$

where for any symmetric matrix $\mathbf{A}$ the notation $\phi_j(\mathbf{A})$ denotes its jth largest eigenvalue.

Under normality Lawley (1956) obtained a Bartlett-correction factor. Fujikoshi (1977) derived an asymptotic expansion of the null distribution of the LR statistic. It may be noted that the normality assumption will be not necessarily satisfied in most applications. So, it is important to examine whether the results under normality can be used for a nonnormal population, or to extend the results under normality to a nonnormal population. The main purpose of this paper is to extend the Bartlett-correction factor in a normal population to an elliptical population, and examine numerical accuracy of our approximations.

It is said for a p-dimensional random vector $\boldsymbol{x}$ to be distributed as a p-dimensional elliptical distribution with location parameter $\boldsymbol{\mu}$ and scale parameter Δ, which is denoted by $E_p(\boldsymbol{\mu}, \Delta)$, if its probability density function has the form

$$c_p|\Delta|^{-1/2}g((\boldsymbol{x}-\boldsymbol{\mu})'\Delta^{-1}(\boldsymbol{x}-\boldsymbol{\mu})),$$

where c_p is a positive constant, Δ is a positive-definite symmetric matrix and $g(\cdot)$ is a nonnegative function (see, e.g., Anderson, 2003). Then the characteristic function has the form

$$C(\boldsymbol{t}) = \exp[\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu}]\psi(\boldsymbol{t}'\Delta\boldsymbol{t})$$

for some function ψ. Note that the expectation and the covariance matrix of $\boldsymbol{x}$ are $E(\boldsymbol{x})=\boldsymbol{\mu}$ and $Cov(\boldsymbol{x})=\boldsymbol{\Sigma}=-2\psi'(0)\Delta$, respectively. Let

$$\kappa = \frac{\psi^{(2)}(0)}{(\psi'(0))^2} - 1 \quad \text{and} \quad \phi = \frac{\psi^{(3)}(0)}{(\psi'(0))^3} - 1. \tag{4}$$

The parameter κ is called the kurtosis parameter.

When we consider the null distribution of T_{LR}, it is no loss of generality that Σ may be diagonal, since the eigenvalues of S are the same as ones of $Q'SQ$ for any orthogonal matrix Q. Furthermore, T_{LR} is invariant due to the multiplication of S by a positive scalar. Therefore, under H_0, we may assume that $\Sigma=\mathrm{diag}(\lambda_1/\lambda, \ldots, \lambda_q/\lambda, 1, \ldots, 1)$. For a notational simplicity, we write

$$\Sigma = \mathrm{diag}(\lambda_1, \ldots, \lambda_q, 1, \ldots, 1). \tag{5}$$

Here λ_i should be read as $\widetilde{\lambda}_i = \lambda_i/\lambda$.

In Section 2, first we derive the limiting distribution of the LR statistic under an elliptical population when p is fixed and N is large. Next we obtain a Bartlett-correction factor, which can be determined by finding the expectation of T_{LR}. In Section 3, we give some empirical results to confirm the effectiveness of our new approximations and compare them with the approximations under normality.

2. The distribution of LR statistic under an elliptical population

2.1. Preliminaries

Lawley (1956) and Fujikoshi (1977) used a stochastic expansion of T_{LR} under a normal population. First, we note that the expansion can be used under a more general population as well as a normal population. Let

$$\mathbf{V} = \sqrt{n}(\mathbf{S}-\boldsymbol{\Sigma}) = \begin{pmatrix}\mathbf{V}_{11} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{pmatrix}, \tag{6}$$