

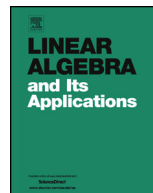


ELSEVIER

Contents lists available at ScienceDirect

Linear Algebra and its Applications

[www.elsevier.com/locate/laa](http://www.elsevier.com/locate/laa)



# Asymptotic properties of a correlation matrix under a two-step monotone incomplete sample



Shin-ichi Tsukada

*School of Education, Meisei University, 2-1-1 Hodokubo, Hino, Tokyo, 191-8506, Japan*

## ARTICLE INFO

### *Article history:*

Received 8 February 2015

Accepted 18 September 2015

Available online 30 September 2015

Submitted by R. Brualdi

### *MSC:*

62H10

62E20

### *Keywords:*

Correlation matrix

Asymptotic distribution

Monotone incomplete sample

## ABSTRACT

This paper derives an asymptotic distribution for a correlation matrix in the context of a two-step monotone incomplete sample drawn from  $N_{p+q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , a multivariate normal population with mean  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . Based on the result, we perform hypothesis testing for the correlation matrix and investigate its accuracy using numerical simulations.

© 2015 Elsevier Inc. All rights reserved.

## 1. Introduction

In multivariate analysis, one of basic matrices is a correlation matrix. In literature concerning the correlation matrix, Neudecker and Wesselman [11] derived an asymptotic distribution for normally distributed observations. Neudecker [10] obtained asymptotic distributions under non-normal distributional assumptions, and Kollo and Ruul [7] derived multivariate density expansions. Furthermore, Aitkin et al. [1] discussed hypothesis testing for the correlation matrix.

*E-mail address:* [tsukada@ed.meisei-u.ac.jp](mailto:tsukada@ed.meisei-u.ac.jp).

In practical data analysis, missing data often appear. Therefore, methods that can use missing data should be employed. These methods have been studied by many authors, including Anderson and Olkin [2], Srivastava [13], Little and Rubin [8], Kanda and Fujikoshi [5], and Chang and Richards [3,4].

In this paper, the asymptotic distribution for the correlation matrix is considered in the context of a two-step monotone incomplete sample drawn from  $N_{p+q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , which is a  $(p + q)$ -dimensional multivariate normal population with mean  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . By deriving the asymptotic distribution for the correlation matrix, we consider its estimation and can conduct hypothesis testing. We suppose the data are composed of  $N$  mutually independent observations consisting of a random sample of  $n$  complete observations and  $N - n$  additional observations on  $\boldsymbol{x}$  alone. That is,

$$\begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{y}_1 \end{pmatrix}, \begin{pmatrix} \boldsymbol{x}_2 \\ \boldsymbol{y}_2 \end{pmatrix}, \dots, \begin{pmatrix} \boldsymbol{x}_n \\ \boldsymbol{y}_n \end{pmatrix}, \begin{pmatrix} \boldsymbol{x}_{n+1} \\ * \end{pmatrix}, \dots, \begin{pmatrix} \boldsymbol{x}_N \\ * \end{pmatrix} \sim N_{p+q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{1}$$

where  $\boldsymbol{x}$  is a  $p \times 1$  vector,  $\boldsymbol{y}$  is a  $q \times 1$  vector, and the symbol  $*$  denotes the missing data. The data in (1) are usually referred to as a two-step monotone incomplete sample and show the simplest pattern available with the missing data. The maximum likelihood estimators (MLEs) of  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  can be explicitly expressed and are given by Anderson and Olkin [2]. Properties of the MLE are discussed by Kanda and Fujikoshi [5] and Chang and Richards [3,4]. Tsukada [14] proposed an unbiased estimator for the covariance matrix  $\boldsymbol{\Sigma}$ . Recently, studies for this sample have been actively conducted, such as Shutoh [12], Tsukada [14–16], Yamada et al. [17], and Yamada [18].

The remainder of the paper is structured as follows. Preliminary results are described in Section 2. Section 3 discusses the asymptotic properties for the correlation matrix. The accuracy of these results is investigated by performing numerical simulations in Section 4. Finally, Section 5 contains a brief summary and ends with a discussion on possible directions for future research.

## 2. Preliminary results

Assume observation (1) is available. The complete samples  $(\boldsymbol{x}'_i, \boldsymbol{y}'_i)'$ ,  $i = 1, \dots, n$  are drawn from  $N_{p+q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , a multivariate normal population with mean  $\boldsymbol{\mu} = (\boldsymbol{\mu}'_1, \boldsymbol{\mu}'_2)'$  and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

whereas the incomplete samples  $\boldsymbol{x}_i$ ,  $i = n + 1, \dots, N$  are drawn from  $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ . It is assumed that all  $N$  samples are mutually independent, and the data are missing at random (MAR) to ignore the missingness mechanism. Lu and Copas [9] noted that inference using the likelihood method is valid if and only if the missing data mechanism is MAR.

Download English Version:

<https://daneshyari.com/en/article/4598815>

Download Persian Version:

<https://daneshyari.com/article/4598815>

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