Available online at www.sciencedirect.com
SCIENCE DIRECT®

LINEAR ALGEBRA AND ITS APPLICATIONS

Linear Algebra and its Applications 410 (2005) 143–149
www.elsevier.com/locate/laa

A multivariate version of Samuelson's inequality

Götz Trenkler [a,*], Simo Puntanen [b]

[a] Department of Statistics, University of Dortmund, Vogelpothsweg 87, D-44221 Dortmund, Germany
[b] Department of Mathematics, Statistics and Philosophy, University of Tampere, FI-33014, Finland

Received 10 January 2003; accepted 21 February 2005
Available online 27 April 2005
Submitted by G.P.H. Styan

Abstract

According to Samuelson's inequality, the deviation of any particular observation from the mean is bounded by a multiple of standard deviation. In this paper we consider some properties of a multivariate extension of this result.

AMS classification: 62H20; 15A45

Keywords: Covariance matrix; Mahalanobis distance; Matrix inequalities; Orthogonal projector; Sample mean; Sample variance

1. Introduction

It is well known that for a real data set $x_1, x_2, \ldots, x_n$, the Samuelson's inequality

$$(x_j - \bar{x})^2 \leqslant (n-1)s^2 \tag{1.1}$$

holds for all $j = 1, \ldots, n$, where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ denotes the mean and $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is the sample variance; see [6,10,14]. Actually, this result can be traced back to Thompson [12] and even earlier to Laguerre [8]. We may cite Olkin [9, p. 205] who states that although the inequality (1.1) appeared in early literature,

* Corresponding author.
E-mail addresses: trenkler@statistik.uni-dortmund.de (G. Trenkler), simo.puntanen@uta.fi (S. Puntanen).

0024-3795/$ - see front matter
doi:10.1016/j.laa.2005.02.032

it was not popularly known until the article by Samuelson [10]. Excellent overviews of the development or Samuelson's inequality can be found in [7,9].

The relationship (1.1) is sharper than the trivial inequality

$$(x_j - \bar{x})^2 \leqslant ns^2. \tag{1.2}$$

Equality occurs in (1.1) if and only if all the x_i other than x_j are equal, whereas in (1.2) we have equality if and only if all x_i are identical. In the following we will consider a multivariate generalization of Samuelson's inequality, of which one form was given by Arnold and Groeneveld [1]. We offer a new proof, based on specific updating formulas, and in particular, offer an equivalent way to express the multivariate case in terms of nonnegative definiteness.

2. Extension of Samuelson's inequality

Suppose that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represents a random sample from a p-dimensional distribution with the corresponding $n \times p$ observation matrix

$$\mathbf{X}_n = (\mathbf{x}_1, \ldots, \mathbf{x}_n)' = (x_{ij}). \tag{2.1}$$

Then the mean of the jth variable is given by $\bar{x}_j = \frac{1}{n}\sum_{i=1}^n x_{ij}$, and the vector of all p means is

$$\bar{\mathbf{x}}_n = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i = (\bar{x}_1, \ldots, \bar{x}_p)' = \frac{1}{n}\mathbf{X}_n'\mathbf{1}_n, \tag{2.2}$$

where $\mathbf{1}_n$ is the $n \times 1$ vector of ones. Consider now the $p \times p$ sample covariance matrix $\mathbf{S}_n = (s_{ij})$ with entries

$$s_{ij} = \frac{1}{n}\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j). \tag{2.3}$$

In matrix notation we have

$$\mathbf{S}_n = \frac{1}{n}\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}_n)(\mathbf{x}_i - \bar{\mathbf{x}}_n)' = \frac{1}{n}\mathbf{X}_n'\mathbf{C}_n\mathbf{X}_n, \tag{2.4}$$

where $\mathbf{C}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n' = \mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n}$ is the so-called centering matrix; here $\mathbf{P}_{\mathbf{1}_n}$ refers to the orthogonal projector (with respect to the standard inner product) onto the column space of matrix (vector) $\mathbf{1}_n$.

In the following, we assume that $n > p$ and for all n the sample covariance matrix $\mathbf{S}_n$ is nonsingular. For singularity of $\mathbf{S}_n$ we refer to [13].

Following [3,11], consider weights w_i associated with the observation vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and define the following quantities:

$$W_n = \sum_{i=1}^n w_i, \quad \bar{\mathbf{z}}_n = \frac{1}{W_n}\sum_{i=1}^n w_i\mathbf{x}_i, \quad \mathbf{Q}_n = \sum_{i=1}^n w_i(\mathbf{x}_i - \bar{\mathbf{z}}_n)(\mathbf{x}_i - \bar{\mathbf{z}}_n)', \tag{2.5}$$