Appl. Statist. (2004)
53, *Part* 4, *pp.* 539–555

Simple component analysis

Valentin Rousson and Theo Gasser

University of Zürich, Switzerland

[Received June 2003. Revised September 2003]

Summary. With a large number of variables measuring different aspects of a same theme, we would like to summarize the information in a limited number of components, i.e. linear combinations of the original variables. Among linear dimension reduction techniques, principal component analysis is optimal in at least two ways: principal components extract the maximum of the variability of the original variables, and they are uncorrelated. Unfortunately, they are often difficult to interpret. Moreover, in most applications, only the first principal component is a 'block component', the remaining components being 'difference components' which are also more difficult to interpret. The goal of simple component analysis is to replace (or to supplement) principal components with suboptimal but better interpretable 'simple components'. We propose a fast algorithm which seeks the optimal system of components under constraints of simplicity. Thus, in contrast with other techniques like 'varimax', this approach always provides a simple solution. The optimal simple system is suboptimal compared with principal components: less variability is extracted and components are correlated. However, if the loss of extracted variability is small, and correlations between components are low, it might be advantageous for practical use. Moreover, our concept of simplicity allows the system to have more than one block component, which also facilitates interpretation. Simplicity is not a guarantee for interpretability. With the help of our algorithm, the user can partly modify an optimal simple system of components to enhance interpretability. In this respect, the ultimate goal of simple component analysis is not to propose a method that leads automatically to a unique solution, but rather to develop tools for assisting the user in his or her choice of an interpretable solution. Finally, we argue that simple components may also make the task of choosing the dimension easier. The methodology is illustrated with a test battery to study the development of neuromotor functions in children and adolescents.

Keywords: Block structure; Classification of variables; Correlation structure; Interpretability of components; Principal components; Reduction of dimensionality; Shrinkage methods

1. Introduction

Principal component analysis (PCA) is an efficient dimension reduction technique which allows us to summarize p variables in q components, where q can be considerably smaller than p. Principal components are defined as optimal linear combinations of the original variables, extracting a maximum of variability and being uncorrelated. Unfortunately, the interpretation of principal components is in general neither easy nor straightforward (see proposals by Jeffers (1967), Jackson (1991) and Cadima and Jolliffe (1995)).

A test battery for assessing the development of neuromotor functions in children and adolescents is used as an example. It consists of 20 standardized tasks of different complexity for which timed performance is measured. These 20 tasks are listed in Table 1. The first 16 tasks are fine motor tasks. They involve either the feet (tasks 1–4), the hands (tasks 5–8) or the fingers (tasks 9–12). Tasks 13–16 are pegboard tasks, where small pieces of metal must be put into holes. Tasks

Address for correspondence: Valentin Rousson, Department of Biostatistics, Institute for Social and Preventive Medicine, University of Zürich, Sumatrastrasse 30, CH-8006 Zürich, Switzerland.
E-mail: rousson@ifspm.unizh.ch

 0035–9254/04/53539