Selection of Components in Principal Component Analysis as a Model Selection Method

Yasunori Fujikoshi
Department of Mathematics
Faculty of Science and Engineering
Chuo University
Kasuga, Bunkyo-ku, 112-8551, Japan

ABSTRACT: Many methods have been proposed to determine the number of relivant components in principal component analysis (see, e.g. Jolliffe (1988), Ferré (1995)). The methods will depend on the aims of statistical analysis and the models considered. In this paper we consider two types of models. Let S be the sample covariance matrix based on a random sample of $N = n+1$ from a p-dimensional population with mean vector μ and covariance matrix Σ. The first one is to select an appropriate model from the set of models $M_k : \lambda_{k+1} = \cdots = \lambda_p, (k = 0, 1, \ldots, p-1)$, where $\lambda_1 \geq \cdots \geq \lambda_p$ are the characteristic roots of Σ. We give a selection criterion by applying Akaike's idea (Akaike (1973)) for a closeness measure $D(S, \Sigma) = -n \log |\Sigma^{-1} S| + n\mathrm{tr}\Sigma^{-1}(S - \Sigma)$ between S and Σ. Its bootstrap version is also given. Next we reexamine another class of models. Let $Y = (y_{ij})$ be the $N \times p$ observation matrix. Consider a class of models $Y = \mathbf{1}_N \mu' + \sum_{j=1}^{k} \sqrt{\lambda_j} \beta_j \gamma_j' + \varepsilon, (k = 0, 1, \ldots, p-1)$, where γ_j's are parameter vectors satisfying $\gamma_i' \gamma_j = \delta_{ij}$, and the elements of the error matrix ε are independently distributed as $N(0, \sigma^2)$. The β_j's are parameter vectors satisfying $\beta_i' \beta_j = \delta_{ij}$ or random vectors whether its model is fixed or random. After reviewing some works, we propose a model selection criterion when the random effects β_j's are given. The criteria in a high-dimensional situation are also examined through a simulation study, etc.

Keywords: *Estimation of dimensionality*, *model selection method*, and *principal componet analysis*

References

AKAIKE, H. (1973): Informaiton theory and an extension of the maximum likelihood principle. In B. N. Petrov and F.Csáki (Eds.): *2nd International Symposium on Information Theory*, Akadémia Kiado, Budapest, 267-281.

FERRÉ, L. (1995): Selection of componets in principal component analysis: A comparison of methods. *Computational Statistics & Data Analysis, 19, 669-682.*

JOLLIFFE, I.T. (1988): Principal Component Analysis. *Springer Verlag.*