On Likelihood Ratio Tests for Dimensionality Selection

Yoshio Takane[1], Peter G. M. van der Heijden[2], and Michael W. Browne[3]

[1] Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, QC H3A 1B1 Canada. *takane@takane2.psych.mcgill.ca*

[2] Department of Methodology and Statistics, Utrecht University, 3508 TC Utrecht, The Netherlands. *p.vanderheijden@fsw.ruu.nl*

[3] Department of Psychology, Ohio State University, 1885 Neil Avenue Mall, Columbus, OH 43210-1222 USA. *mbrowne@magnus.acs.ohio-state.edu*

Summary. It has been pointed out that parameterized mixture models (e.g., normal mixtures, multi-layered neural networks, etc.) have points of irregularity in their parameter space, which in some cases invalidates the use of likelihood ratio (LR) tests. We point out that the problem is much more ubiquitous. Virtually all models requiring dimensionality selection (the number of clusters, latent classes, canonical variates, factors, components, dimensions, etc.) have similar problems. In this paper we investigate the nature of the problem both theoretically and by Monte-Carlo studies.

Key words: Models with dimensional structure, Local regularity, Asymptotic chi-square distribution.

1. Introduction

Many multivariate statistical models have dimensional structure. Such models typically require judicious choice of dimensionality. In multidimensional scaling (MDS), for example, similarity data between stimuli are represented by distances between points in a multidimensional space. A natural question that arises is how many dimensions are necessary to approximate the observed similarity data. Models with similar structure include latent class analysis, factor analysis, principal component analysis, canonical correlation analysis, reduced-rank regression, RC-association and correlation models, correspondence analysis, and so on.

Likelihood ratio (LR) tests are often used for dimensionality selection. However, there is a great deal of confusion regarding the asymptotic distribution of the LR statistics. Although the asymptotic distribution of the LR statistic (say, λ_s) representing the difference between the correct model of minimal dimensionality and the saturated model is indeed chi-square, that of the LR statistic (say, λ_c) representing the difference between the correct model and the one with one dimensional higher than the correct model is not likely to be chi-square due to a violation of one of the regularity conditions. In this paper we attempt to clarify the misunderstanding. This common misunderstanding has occurred repeatedly in various fields.

2. The Condition for Local Regularity

Let $\boldsymbol{\theta}$ be the parameter vector, and let $\boldsymbol{\xi}$ be the vector of model predictions, which is a function of $\boldsymbol{\theta}$, i.e., $\boldsymbol{\xi} = \boldsymbol{g}(\boldsymbol{\theta})$ for some analytic function $\boldsymbol{g}$. Let $\boldsymbol{J}(\boldsymbol{\theta})$ be the Jacobian matrix defined by $\boldsymbol{J}(\boldsymbol{\theta}) = [\partial \boldsymbol{g}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}']$.

Local Regularity. A point $\boldsymbol{\theta}_0$ in the interior of the parameter space $\boldsymbol{\Theta}$ is said to be locally regular if $\text{rank}(\boldsymbol{J}(\boldsymbol{\theta})) = \text{rank}(\boldsymbol{J}(\boldsymbol{\theta}_0))$ for all $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$.

We illustrate that this local regularity condition is violated for comparison between the r-dimensional correct model and the $(r+1)$-dimensional model. Assume, for simplicity, that there are four points in the two-dimensional Euclidean space. Let x_{it} denote the coordinate of point i on dimension t. We set $x_{11} = x_{12} = 0$ to remove the translational indeterminacy, and $x_{22} = 0$ to remove the rotational indeterminacy in the Euclidean space. Thus, the parameter vector is: $\boldsymbol{\theta} = (x_{21}, x_{31}, x_{41}, x_{32}, x_{42})'$, and the model vector is: $\boldsymbol{\xi} = (d_{12}, d_{13}, d_{14}, d_{23}, d_{24}, d_{34})'$, where $d_{ij} = \{\sum_{t=1}^{2}(x_{it} - x_{jt})^2\}^{1/2}$. The Jacobian matrix for this model is given by:

$$\boldsymbol{J}(\boldsymbol{\theta}) = \begin{bmatrix} -\delta_{121} & 0 & 0 & 0 & 0 \\ 0 & -\delta_{131} & 0 & -\delta_{132} & 0 \\ 0 & 0 & -\delta_{141} & 0 & -\delta_{142} \\ \delta_{231} & -\delta_{231} & 0 & -\delta_{232} & 0 \\ \delta_{121} & 0 & -\delta_{241} & 0 & -\delta_{242} \\ 0 & \delta_{341} & -\delta_{341} & \delta_{342} & -\delta_{342} \end{bmatrix},$$

where $\delta_{ijt} = (x_{it} - x_{jt})/d_{ij}$. It is assumed that $d_{ij} \neq 0$. The rank of this Jacobian matrix is 5 for the two-dimensional model. For the one-dimensional model, $x_{32} = x_{42} = 0$, so that the rank of the Jacobian matrix reduces to 3. This shows that point $\boldsymbol{\theta}_0$ is not locally regular, since arbitrary perturbations of $x_{32} = 0$ and $x_{42} = 0$, however small they may be, increase the rank of the Jacobian matrix. Under this circumstance, the asymptotic distribution of the LR statistic (λ_c) for representing the difference between the one-dimensional model (assumed correct) and the two-dimensional model is not necessarily chi-square. Similar observations can be made for other models with similar dimensional structure. Many of the models mentioned above, however, have a saturated model. The asymptotic distribution of the LR statistic (λ_s) representing the difference between the r-dimensional model and the saturated model is still chi-square, if the r-dimensional model is correct but the $(r-1)$-dimensional model is not correct. This statistic may be used for dimensionality selection in these models. A lot of confusion in the literature stems from the misunderstanding that the asymptotic distribution of λ_c is also chi-square like λ_s.

For models with no saturated model (e.g., normal mixtures, multi-layered neural networks, reduced-rank regression, etc.), the situation is considerably more complicated. In this case, λ_s cannot be defined, and λ_c is not likely to follow an asymptotic chi-square distribution. Some attempts are being made to find the order of convergence of this statistic.

In Higuchi, T., Iba, Y., and Ishiguro, M. (Eds.), *Proceedings of Science of Modeling: The 30th Anniversary Meeting of the Information Criterion (AIC)*, (pp. 348-349). Tokyo: The Institute of Statistical Mathematics, 2003.