Robust Diagnostics in Graphical Modeling

Woosung Yang
Okayama University, Graduate School of Natural Science and Technology
3-1-1, Tsushima-Naka
Okayama, Japan
yangws@ems.oakayama-u.ac.jp

Yutaka Tanaka
Okayama University, Faculty of Environmental Science & Technology
3-1-1, Tshusima-Naka
Okayama, Japan
tanaka@ems.okayama-u.ac.jp

1. Introduction

For accessing jointly as well as individually influential observations in graphical modeling Yang, Sung and Tanaka(2001) propose a method based on Cook's local influence(Cook, 1986). However, as their method uses the MLEs of the mean vector and the covariance matrix in their diagnostic statistics, there is a possibility that their method suffers from the so-called masking effect. Recently, Atkinson and Riani(2000) proposed an idea of forward search in regression analysis to protect from suffering the masking effect. Their idea can be applied to other multivariate methods. In the present paper we try to combine those two methods to detect influential observation without suffering from the masking effect.

2. Local influence in graphical modeling

Here we consider graphical Gaussian models or covariance selection models. Let $\underline{\theta} = vech(\Sigma)$, $\underline{\varphi}$ be partial correlations which we are interested in, and $L(\underline{\theta})$ be the likelihood function. Then the normal curvature is given by

$$C_d(\underline{\varphi}) = 2\left|\underline{d}^T\left(\frac{\partial \underline{\varphi}^T}{\partial \underline{w}}\right)\left[\frac{\partial \underline{\varphi}}{\partial \underline{\theta}^T}(-\ddot{L})^{-1}\frac{\partial \underline{\varphi}^T}{\partial \underline{\theta}}\right]^{-}\left(\frac{\partial \underline{\varphi}}{\partial \underline{w}^T}\right)\underline{d}\right|,$$

where it is assumed that the perturbation parameters $\underline{w}$ change from $\underline{w}_0$ to $\underline{w}_0 + t\underline{d}$ along a direction $\underline{d}$. In Cook's local influence dominant eigenvalues and the associated eigenvectors of the coefficient matrix of the above quadratic form of $\underline{d}$ play important roles.

3. Forward search

The basic idea of forward search is to order the observations by their closeness to the model. The starting point is found by fitting to a large number of small subsets, using methods from robust statistics to determine which subset fits best. We then order all observations by their closeness to this model. The forward search algorithm is made up of three steps: the first concerns the choice of an initial subset, the second refers to the way in which we progress in the forward search and the third relates to the monitoring of the statistics during the progress of the search. We start with a set of p observations, which is chosen in the first step in a robust manner. Using the sample mean vector and sample covariance matrix of the chosen observations we calculate the Mahalanobis distance(MD_i) for all observations, and then in the next step choose the $p+1$ observations with smallest MD_i. The forward search is repeated in this way until all observations are chosen in the subset, and at each step some diagnostic statistics are monitored.

4. Our procedure

In monitoring the forward search we additionally calculate Cook's D_i, where D_i is defined as C_d with $\underline{d} = (d_1, \cdots, d_n)^T,\ d_j = \delta_{ij},\ j = 1, \cdots, n$, δ_{ij} indicating the Kronecker's δ. We study numerically how $\{D_i,\ i = 1, \cdots, n\}$ change in the forward search process.

REFERENCES

Atkinson A. and Riani, M. (2000). *Robust Diagnostic Regression Analysis. Springer.*

Cook, R. D. (1986). Assessment of local influence. *J. R. Statist. Soc.*, **B48**, 133-169.

Yang, W. S., Sung, J. M. and Tanaka, Y. (2001). Local influence in graphical modeling. *ICNCB 2001 Proceedings*

RÉSUMÉ

On propose une méthode robuste pour détecter des observations influentes dans modeler graphique. L'idée fondamentale est de se protéger contre le prétendu effet masquant en combinant Atkinson et Riani(2000) recherchent en avant et des influences locales de Cook(1996). Un exemple numérique est donné pour montrer l'exécution de la méthode proposée.