

An individual influence using resampling method and its evaluation

Masaya Iizuka

Okayama University, Faculty of Law

3-1-1, Tsushima naka

Okayama 700-8530, Japan

masa@law.okayama-u.ac.jp

Yuichi Mori

Okayama University of Science, Department of Socio-Information

1-1, Ridaityou

Okayama 700-0005, Japan

mori@soci.ous.ac.jp

Yutaka Tanaka and Tomoyuki Tarumi

Okayama University, Department of Environmental and Mathematical Sciences

3-1-1, Tsushima naka

Okayama 700-8530, Japan

tanaka@ems.okayama-u.ac.jp, t2@ems.okayama-u.ac.jp

1. Introduction

In the resampling method, bootstrap diagnostics was proposed recently. Jackknife Influence Function(JIF) and relative JIF are used as an indicator which measures influence in the method, and the main roles when examining the sensitivity of each observations. On the other hand, sensitivity analysis in multivariate analysis, the method based on Influence Function (IF) and the method based on Cook's Local Influence (LI) are mainly used. Tanaka (1994) is discussed both relation. Difference is that whether a difference of the method based on IF and bootstrap diagnostics is approximates by differentiation or actually deleted. To evaluate the result in principle component analysis, we compared the result applied to the sensitivity analysis based on an influence function with the result applied to jackknife-after-bootstrap method.

2. Influence Function

Empirical Influence Function(EIF)

Influence function of i -th observation is defined by

$$EIF(\mathbf{x}_i; \hat{\theta}) \equiv \lim_{\epsilon \rightarrow 0} \left\{ \theta \left[(1 - \epsilon)\hat{F} + \epsilon\delta(\mathbf{x}) \right] - \theta(\hat{F}) \right\} / \epsilon.$$

Let's the parameter of the i -th observation deleted denote by $\hat{\theta}_{(i)}$. Using influence function, we obtain the following; $\hat{\theta}_{(i)} \cong \hat{\theta} - (n - 1)^{-1}EIF(\mathbf{x}_i; \hat{\theta})$. We evaluate the influence of the i -th observation by this value.

Jackknife influence function

We apply jackknife-after-bootstrap in resampling method which is introduced by Efron (1992). Jackknife-after-bootstrap method uses only B bootstrap samples to obtain the information.

Let $\hat{\gamma}_{(i)}$ is bootstrap statistic. Jackknife influence function for $\hat{\gamma}_{(i)}$ is defined by

$\tilde{u}_i\{\hat{\gamma}\} = (n - 1)(\tilde{\gamma}_{(i)} - \tilde{\gamma}_{(\cdot)})$, where $\tilde{\gamma}_{(\cdot)} = \sum_i \hat{\gamma}_{(i)}/n$, and relative jackknife influence function is denoted by $\tilde{u}_i^*\{\hat{\gamma}\} = \tilde{u}_i\{\hat{\gamma}\} / \left(\sum_j \frac{\tilde{u}_j\{\hat{\gamma}\}^2}{n-1}\right)^{\frac{1}{2}}$. Thus, we compare the value that obtained by JIF and relative JIF with the value of EIF.

3. evaluation

To evaluate these method, for example, we calculate Cook's D of eigenvalue using both method.

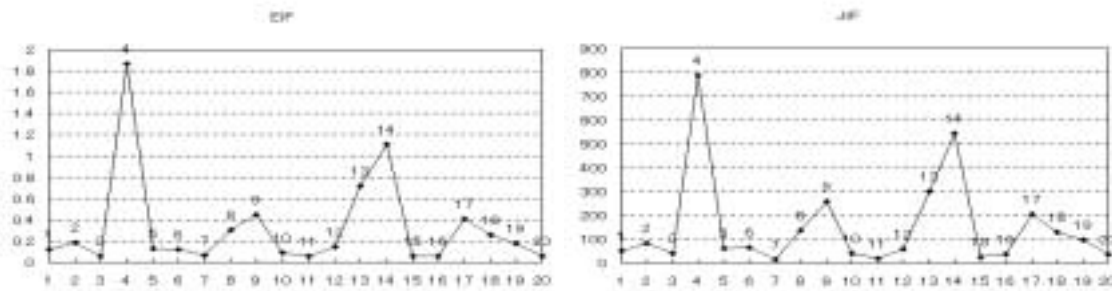


Figure 1. index plot(eigenvalue):Cook's D -number of PCs 2-

REFERENCES

- Bradley Efron. (1992). Jackknife-after-Bootstrap Standard Errors and Influence Functions, *J. R. Statist. Soc. B.* **54**, 83-127.
- Bradley Efron, Robert J. Tibshirani (1993). *An Introduction to the Bootstrap*. Chapman & Hall.
- Tanaka, Y. (1994). Recent advance in sensitivity analysis in multivariate methods. *J. Jpn. Soc. Comp. Statist.*, **7**, 1-25.

RÉSUMÉ

La manière de se rapprocher par différenciation on celle d'effacer les données réellement, c'est une différence entre la méthode basée en fonction IF et le diagnostic d'amorce. Pour évaluer le résultat en principe l'analyse composante, nous avons comparé le résultat appliqué à l'analyse de sensibilité basée sur une fonction d'influence avec le résultat appliqué à la méthode d'jackknife-apres-bootstrap.