Journal of the Korean Statistical Society 43 (2014) 1–9

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

Sensitivity analysis for ranked data☆

Sang Tae Han *

Department of Informational Statistics and Basic Science, Institute of Hoseo University, Asan 336-795, Republic of Korea

ARTICLE INFO

Article history:
Received 23 April 2013
Accepted 22 November 2013
Available online 27 December 2013

AMS 2000 subject classifications:
primary 62H25
secondary 62P25

Keywords:
Sensitivity analysis
Singular value decomposition
Ranked data analysis
Perturbation theory
Influence derivatives

ABSTRACT

Sensitivity analysis is to study the influence of a small change in the input data on the output of the analysis. Han and Huh (1995) developed a quantification method for the ranked data. However, the question of stability in the analysis of ranked data has not been considered. Here, we propose a method of sensitivity analysis for ranked data. Our aim is to evaluate perturbations by using a graphical approach suggested by Han and Huh (1995). It extends the results obtained by Tanaka (1984) and Huh (1989) for the sensitivity analysis in Hayashi's third method of quantification and those by Huh and Park (1990) for the principal component reduction of the case influence derivatives in regression. A numerical example is provided to explain how to conduct sensitivity analysis based on the proposed approach.

1. Introduction

The aim of the sensitivity analysis is to evaluate the influence on the result of analysis caused by small changes in the input data. In linear regression analysis, numerous methods have been proposed for the analysis of the influence of a single or multiple observations on the results of analysis and major parts of them have been summarized in several books; Belsley, Kuh, and Welsch (1980) and Cook and Weisberg (1982) among others. The underlying idea is not restricted to the regression analysis but is commonly applied to other statistical methods including multivariate methods; see Critchley (1985), Huh (1989) and Tanaka (1984, 1994). Specially, Critchley (1985) has considered the influence on the eigenvalues and eigenvectors as the major results in principal component analysis and they have evaluated the influence of each observation on the eigenvalues.

However, sensitivity analysis has not been developed for ranked data. If configurations (quantification results of judges and objects) depend heavily upon a few judges, we must be very careful with using the result. Therefore, it may be worth to find out whether the configuration of the judges or objects is stable or not. To investigate such a problem, we propose a method of sensitivity analysis for ranked data based on the perturbation theory of eigenvalues and eigenvectors. The major aim of the proposed method is to evaluate the influence of data perturbation (such as case deletions) on the results of analysis. For convenience, the method of quantification for ranked data in Han and Huh (1995) is briefly explained.

Let r_{ij} denote the rank given to the object $j(=1, \ldots, p)$ by the judge $i(=1, \ldots, n)$ and write $R = \{r_{ij}\}$. As a row centering process, let

$$s_{ij} = r_{ij} - (p+1)/2$$

☆ This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government [NRF-2011-0022809]. This research was studied during the sabbatical year sponsored by Hoseo University.

* Tel.: +82 415405901; fax: +82 415405908.
E-mail addresses: sthan@hoseo.edu, sthan4@naver.com.

1226-3192/$ – see front matter © 2013 The Korean Statistical Society. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jkss.2013.11.004

and write

$$S = \{s_{ij}\}.$$

In Spearman's sense, the squared rank distance between two rows i and i'(=1, 2, ..., n) is defined as

$$d_S^2(i, i') = \sum_{j=1}^{p}(s_{ij} - s_{i'j})^2.$$

Then, let s_i be the ith row of S in $\Re^p$, and let υ be the unit vector $\Re^p$. The magnitude of the projection vector of s_i on υ is equal to $s_i'\upsilon$. Consider the following optimization problem:

$$\max_{\upsilon} \sum_{i=1}^{n}(s_i'\upsilon)^2, \quad \text{subject to } \upsilon'\upsilon = 1.$$

Then, the solution is related to the singular value decomposition of S. Let

$$S = UDV',$$

where U is a $n \times p$ matrix with orthonormal columns, V is a $p \times p$ orthonormal matrix, and D is the $p \times p$ diagonal matrix with singular values, $\lambda_1^{1/2} \geq \cdots \geq \lambda_{p-1}^{1/2} \geq \lambda_p^{1/2}$ as its diagonal elements. Here $\lambda_p^{1/2} = 0$ (because, rank$(S) = p - 1$), and therefore write

$$S = U_1D_1V_1',$$

where U_1 is a $n \times (p-1)$ submatrix of $U(U_1'U_1 = I_{p-1})$, V_1 is the $p \times (p-1)$ submatrix of $V(V_1'V_1 = I_{p-1})$, and D_1 is the $(p-1) \times (p-1)$ diagonal submatrix of D. Then the plot points, x_i ($i = 1, \ldots, n$) and y_j ($j = 1, \ldots, p$), are given by the row vectors of X and Y, where

$$X = SV_1 = U_1D_1,$$
$$Y = V_1.$$

For further details, see Han and Huh (1995).

The rest of this study is organized as follows: Section 2 develops the perturbation theory of the eigenvalue–eigenvector problem for ranked data. Section 3 proposes the principal component reduction of case influence derivatives for ranked data. The proposed method enables the identification of subgroups of influential observations (or judges) with visible directions of influence. Section 4, presents a numerical example to demonstrate the usefulness of the proposed method, and in Section 5, conclusions and discussions are given.

2. Perturbation theory for sensitivity analysis

Tanaka (1984) studied the perturbation problem for the case in which rows or columns of a response pattern table are perturbed with infinitesimal change of weights. And, Huh (1989) extends this to a more general case, as elementwise perturbations. The scheme and techniques in Huh's method can also be applied to ranked data in a similar way, as follows.

Suppose that the input data matrix S is perturbed arbitrarily with the derivative $\dot{S}$. More specifically, let S^ϵ denote the perturbed input data matrix satisfying

$$\lim_{\epsilon \to 0} S^\epsilon = S$$

and assume that S^ϵ is continuously differentiable with respect to ϵ. Define

$$\dot{S} = dS/d\epsilon = \lim_{\epsilon \to 0}(S^\epsilon - S)/\epsilon.$$

Other matrix derivatives are defined similarly. Then the derivative $\dot{S}$ of S can be written as

$$\dot{S} = \dot{U}_1D_1V_1' + U_1\dot{D}_1V_1' + U_1D_1\dot{V}_1'.$$

If $H = S'S = V_1\Theta V_1$ is set for $\Theta = D_1^2$, then

$$\dot{H} = \dot{S}'S + S'\dot{S}.$$

In addition, $\dot{H}$ satisfies

$$\dot{H} = \dot{V}_1\Theta V_1' + V_1\dot{\Theta}V_1' + V_1\Theta\dot{V}_1'. \tag{1}$$

Because $V_1'V_1 = I_{p-1}$,

$$\dot{V}_1'V_1 + V_1'\dot{V}_1 = 0. \tag{2}$$