SENSITIVITY ANALYSIS IN HAYASHI'S THIRD METHOD OF QUANTIFICATION

Yutaka Tanaka*

A method of sensitivity analysis is proposed to detect the influential observations in Hayashi's third method of quantification (Hayashi, 1956). It evaluates the changes of the eigenvalues and the scores assigned to the categories and/or individuals due to a small change of the weights for a single or multiple individuals (or categories) by using the perturbation theory of eigenvalue problems.

1. Introduction

Recently many methods have been investigated to evaluate the influence of a single or multiple observations on the result and to detect the influential observations in regression analysis (Belsley, Kuh and Welsch, 1980; Cook and Weisberg, 1982). If the conclusion depends heavily upon one or a few observations, we must be very careful to decide whether to include those observations or not.

The problem of influential observations is not special to regression analysis, but common to the other statistical methods including the methods of quantification. We shall consider the above problem in Hayashi's methods of quantification (Hayashi, 1956). In the present paper we shall take up Hayashi's third method of quantification, which is mathematically equivalent to the correspondence analysis (Benzecri, 1973). We shall use the term sensitivity analysis because various types of data other than ordinary multivariate observations are treated in quantification and the problem is to evaluate how a small change of the input (data) has an influence upon the output (result of analysis).

In section 2 we briefly explain Hayashi's third method of quantification. Then in section 3 we propose a method to evaluate the influence of a single or multiple individual(s) or category(ies) on the result of analysis by using the perturbation theory of eigenvalue problems and in section 4 we give numerical examples to show the usefulness of the proposed method. Finally in section 5 discussion and summary are given.

The fundamental technique of our work is the perturbation theory of eigenvalue problems. The same technique has been used in many statistical fields, for example, in the investigation of the robustness of classical scaling by Sibson (1979), of the asymptotic distributions of statistics based on the sample correlation matrix by Konishi (1979) and of the asymptotic theories of the methods of optimal scaling by Tanaka (1978).

2. Hayashi's third method of quantification

Suppose that the response patterns of \( n \) individuals to \( R \) categories of qualitative attributes are given in the form of Table 1.
Consider the case where we wish to gather together the individuals which respond to the similar categories and also gather together the categories which are responded by the similar individuals. In order to do so we should rearrange the rows and columns to bring the marks close to the diagonal part. In Hayashi’s third method of quantification the rows and columns are quantified operationally so that they can be rearranged to satisfy the above condition by using the assigned scores. For the convenience to develop a method of sensitivity analysis we formulate the method of quantification in a generalized manner.

Now let us assign numerical scores $x_i$ and $y_j$ to the individual $i$ and category $j$, respectively. The scores $\{x_i\}$ and $\{y_j\}$ should be assigned in such a way that the individuals and the categories with similar response patterns have similar values. Then, when the individual $i$ responds to the category $j$, we consider as if we obtain a two-dimensional observation $(x_i, y_j)$ and maximize the correlation coefficient between $\{x_i\}$ and $\{y_j\}$.

Then the problem becomes as follows:

Maximize

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{R} w_i \delta_i(j) x_i y_j,$$

subject to the constraints

$$\sum_{j=1}^{R} g_j y_j / N = 1$$

$$\sum_{i=1}^{n} f_i x_i / N = 1,$$

$$\sum_{j=1}^{R} g_j y_j = 0,$$

$$\sum_{i=1}^{n} f_i x_i = 0,$$

where

$$g_j = \sum_{i=1}^{n} w_i \delta_i(j),$$

$$f_i = \sum_{j=1}^{R} w_i \delta_i(j),$$

$$N = \sum_{i=1}^{n} \sum_{j=1}^{R} w_i \delta_i(j),$$

$$\delta_i(j) = \begin{cases} 1, & \text{if the individual } i \text{ responds to the category } j, \\ 0, & \text{otherwise}, \end{cases}$$

$w_i$ denoting the weight for individual $i$. The constraints (2.2) to (2.5) indicate that the scores $\{x_i\}$ and $\{y_j\}$ have zero means and unit variances.
Using the Lagrange multipliers we obtain the following two eigenproblems of symmetric matrices with common eigenvalues.

\[ \sum_{j=1}^{\gamma} \left( \sum_{i=1}^{n} w_i \delta(i,j) \delta(i,j') \right) (\sqrt{g_j}/N y_j) - \lambda^2 (\sqrt{g_j}/N y_j) = 0, \quad j = 1, 2, \ldots, R, \quad (2.6) \]

\[ \sum_{i=1}^{n} \left( \sum_{j=1}^{\gamma} w_i \delta(i,j) \delta(i,j') \right) (\sqrt{f_i}/N x_i) - \lambda^2 (\sqrt{f_i}/N x_i) = 0, \quad i = 1, 2, \ldots, n, \quad (2.7) \]

The both eigenproblems have a trivial eigenvalue \( \lambda^2 = 1 \) and the associated eigenvectors \((\sqrt{g_j}/N, \ldots, \sqrt{g_n}/N)\) and \((\sqrt{f_1}/N, \ldots, \sqrt{f_n}/N)\), respectively. The other eigenvectors satisfy the constraints (2.4) or (2.5) because they are all orthogonal to the vector \((\sqrt{g_j}/N, \ldots, \sqrt{g_n}/N)\) or \((\sqrt{f_1}/N, \ldots, \sqrt{f_n}/N)\). It can be verified easily that each eigenvalue is equal to the squared correlation coefficient between the corresponding \(y_j\) and \(x_i\). Therefore we should use the eigenvectors \((u_1, \ldots, u_R)\) and \((v_1, \ldots, v_n)\) associated with the largest eigenvalue \(\lambda^2\) excepting \(\lambda^2 = 1\) and assign the scores \(\{y_{i1}\}\) and \(\{x_{i1}\}\) obtained by the following equations.

\[ \begin{align*}
    y_{is} &= u_{is} / \sqrt{g_s}/N, \quad s = 1, 2, \ldots, R, \\
    x_{is} &= v_{is} / \sqrt{f_i}/N, \quad i = 1, 2, \ldots, n,
\end{align*} \quad (2.8) \]

where the eigenvectors \((u_{is})\) and \((v_{is})\) are normalized so that they satisfy \(\|u_{is}\| = \|v_{is}\| = 1\) for any \(s\).

If we cannot grasp the relationship among the categories or individuals well with the above \(\{y_{i1}\}\) and \(\{x_{i1}\}\), we may use the eigenvectors \((u_{i2}, \ldots, u_{R2})\) and \((v_{i2}, \ldots, v_{n2})\) associated with the second largest eigenvalue \(\lambda^2\), and assign the second sets of scores \(\{y_{i2}\}\) and \(\{x_{i2}\}\) obtained by (2.8). These scores are the solution of the maximization problem (2.1) to (2.5) under the additional conditions that \(\{y_j\}\) and \(\{x_i\}\) are uncorrelated with \(\{y_{i1}\}\) and \(\{x_{i1}\}\), respectively. If necessary we may assign the third, fourth, … sets of scores \(\{y_{i3}\}\) and \(\{x_{i3}\}\) calculated in a similar manner.

When either of the set of scores \(\{y_{is}\}\) or \(\{x_{is}\}\) is obtained, the remaining set \(\{x_{is}\}\) or \(\{y_{is}\}\) is calculated by using the following relations.

\[ \begin{align*}
    y_{is} &= \frac{1}{\lambda_s} \sum_{j=1}^{\gamma} w_j \delta(i,j) y_{js}, \quad i = 1, 2, \ldots, n, \\
    x_{is} &= \frac{1}{\lambda_s} \sum_{i=1}^{n} w_i \delta(i,j) x_{is}, \quad j = 1, 2, \ldots, R, \quad (2.9) \end{align*} \]

3. Sensitivity analysis

3.1 Evaluation of the influences of a single observation

By the third method of quantification some \(q\)-dimensional numerical scores \(\{(x_{i1}, \ldots, x_{iq})\}, \quad i = 1, \ldots, n\) and \(\{(y_{j1}, \ldots, y_{jq})\}, \quad j = 1, \ldots, R\) are assigned to the individuals and categories. Using these scores the individuals or categories are expressed as the points in the \(q\)-dimensional Euclidean space and classified into some clusters based on the configurations.

But, if these configurations depend heavily upon a few individuals or categories, we must be very careful in using the result. So it may be worth to know whether the configuration of the categories is stable or not for the selection of individuals, or converse-
ly, whether the configuration of the individuals is stable or not for the selection of categories. In order to investigate such a problem we shall apply the idea of influence function, which is mainly used in regression analysis, to the third method of quantification.

The equations (2.6) and (2.8), which determine the scores \( \{y_j\} \), are rewritten as follows.

\[
\sum_{j=1}^n h_{jj'} u_j - \theta u_j = 0, \quad j = 1, 2, \ldots, R, \tag{3.1}
\]

\[
\sum_{j=1}^n u_j^2 = \frac{\sum_{j=1}^n g_j y_j^2}{N} = 1, \tag{3.2}
\]

where

\[
h_{jj'} = \sum_{i=1}^n \frac{w_i^2 \delta_i(j) \delta_i(j')}{\sqrt{g_j g_{j'}}}, \tag{3.3}
\]

\( \theta \) indicating an eigenvalue \( \lambda^2 \). The equation (3.2) means the normalization that an eigenvector \( \{u_j\} \) has a unit norm and the corresponding scores \( \{y_j\} \) have unit variance. In matrix notations

\[
\begin{bmatrix}
HU &=& U\Theta \\
U'U &=& I
\end{bmatrix}
\tag{3.4}
\]

were \( \Theta \) is a diagonal matrix with the eigenvalues \( \theta_1 = \lambda^2, \theta_2 = \lambda^2, \ldots, \theta_R = \lambda^2 \) in its diagonal parts and \( U \) is a matrix with the eigenvectors in its columns.

Now let the weights for individuals be

\[
w_k = \begin{cases}
1 - \varepsilon, & k = i, \\
1, & k \neq i,
\end{cases} \quad k = 1, 2, \ldots, n. \tag{3.5}
\]

Then the \((j, j')\)-th element \( h_{jj'}(\varepsilon) \) of \( H(\varepsilon) \) is expanded as

\[
h_{jj'}(\varepsilon) = h_{jj'} + \varepsilon h_{jj'}^{(1)} + \varepsilon^2 h_{jj'}^{(2)} + O(\varepsilon^3), \tag{3.6}
\]

where \( h_{jj'} \) is the value of \( h_{jj'}(\varepsilon) \) in the case when \( \varepsilon = 0 \) and

\[
h_{jj'}^{(1)} = -\frac{\delta_i(j) \delta_i(j')}{2 g_j g_{j'}} + \frac{h_{jj'} \left( \delta_i(j) + \delta_i(j') \right)}{g_j g_{j'}},
\]

\[
h_{jj'}^{(2)} = \frac{3}{8} \left( \delta_i(j) + \delta_i(j') \right) + \frac{\delta_i(j) \delta_i(j')}{4 g_j g_{j'}} + \frac{\delta_i(j) \delta_i(j')}{2 g_j g_{j'}} \left( \delta_i(j) + \delta_i(j') \right). \tag{3.7}
\]

Generally, when the matrix \( H(\varepsilon) = (h_{jj'}(\varepsilon)) \) is expanded in a power series convergent in a neighborhood of \( \varepsilon = 0 \), then there exist power series \( \theta_i(\varepsilon), \ldots, \theta_R(\varepsilon) \) and \( u_{i1}(\varepsilon), \ldots, u_{iR}(\varepsilon) \) all convergent in a neighborhood of \( \varepsilon = 0 \) (Kato, 1980; Rellich, 1969). Let \( \Theta + \varepsilon \Theta^{(1)} + \varepsilon^2 \Theta^{(2)} + \ldots \) be the eigenvalues and eigenvectors of \( H(\varepsilon) \), then

\[
\begin{bmatrix}
(H + \varepsilon H^{(1)} + \varepsilon^2 H^{(2)} + \ldots) (U + \varepsilon U^{(1)} + \varepsilon^2 U^{(2)} + \ldots) \\
(\Theta + \varepsilon \Theta^{(1)} + \varepsilon^2 \Theta^{(2)} + \ldots)
\end{bmatrix} = I. \tag{3.9}
\]

Comparing the coefficients of the 1st and 2nd orders of \( \varepsilon \), we obtain the following relations.

1st order:

\[
Z^{(1)} - \Theta Z^{(1)} + \Theta^{(1)} = U' H^{(1)} U, \tag{3.10}
\]

\[
Z^{(1)} + Z^{(1)} = 0. \tag{3.11}
\]
2nd order:
\[
\begin{align*}
Z^{(2)} &= \Theta^{(2)} Z^{(1)} + \Theta^{(1)} Z^{(2)} + \Theta^{(2)} = U^{(1)} H^{(1)} U^{(1)} + U^{(1)} H^{(1)} U^{(1)} - U^{(1)} U^{(1)} \Theta^{(1)} , \\
Z^{(2)} + Z^{(2)} &= -U^{(1)} U^{(1)} , 
\end{align*}
\]  
(3.12)

where
\[
Z^{(1)} = U^{(1)} U^{(1)} , \quad Z^{(2)} = U^{(1)} U^{(1)} .
\]  
(3.13)

If it is assumed that the eigenvalues of interest \(\{\theta_s, s = 1, \ldots, q\}\) are all distinct, the 1st and 2nd order terms of eigenvalues and eigenvectors are expressed as follows.

1st order:
\[
\begin{align*}
\theta_s^{(1)} &= \sum_k \sum_k h_{kk}^{(1)} u_{ks} u_{ks}, \\
Z_{rs}^{(1)} &= \frac{1}{\theta_s - \theta_r} \sum_k h_{kk}^{(1)} u_{kr} u_{ks}, \quad r \neq s, \\
Z_{ss}^{(1)} &= 0, \\
u_{js}^{(1)} &= \sum_r u_{jr} Z_{rs}^{(1)}. 
\end{align*}
\]  
(3.14)

2nd order:
\[
\begin{align*}
\theta_s^{(2)} &= \sum_k \sum_k h_{kk}^{(2)} u_{ks} u_{ks} + \sum_k \sum_k h_{kk}^{(1)} u_{ks} u_{ks} - \theta_s^{(1)} \sum_k u_{ks} u_{ks}, \\
Z_{rs}^{(2)} &= \frac{1}{\theta_s - \theta_r} \left( \sum_k \sum_k h_{kk}^{(2)} u_{kr} u_{ks} + \sum_k \sum_k h_{kk}^{(1)} u_{kr} u_{ks} - \theta_s^{(1)} \sum_k u_{kr} u_{ks} \right), \\
Z_{ss}^{(2)} &= -\frac{1}{2} \sum_k (u_{ks}^{(2)})^2, \\
u_{js}^{(2)} &= \sum_r u_{jr} Z_{rs}^{(2)}. 
\end{align*}
\]  
(3.15)

Hence, from (2.8), (3.14) and (3.15),
\[
\begin{align*}
\theta_s^{(1)} &= \frac{d\theta_s}{de} \bigg|_{e=0} = \sum_k \sum_k h_{kk}^{(1)} u_{ks} u_{ks}, \quad s = 1, \ldots, q, \\
\theta_s^{(2)} &= \frac{1}{2} \frac{d^2\theta_s}{de^2} \bigg|_{e=0} = \sum_k \sum_k h_{kk}^{(2)} u_{ks} u_{ks} + \sum_k \sum_k h_{kk}^{(1)} u_{ks} u_{ks} - \theta_s^{(1)} \sum_k u_{ks} u_{ks}, \\
y_{js}^{(1)} &= \frac{dy_{js}}{de} \bigg|_{e=0} = \sqrt{N g_j} \left( \frac{3 \delta_i(j) - f_i}{2 N g_j} \right) u_{js} + u_{js}^{(1)}, \\
y_{js}^{(2)} &= \frac{1}{2} \frac{d^2y_{js}}{de^2} \bigg|_{e=0} = \sqrt{N} g_j \left( \frac{3 \delta_i(j)}{(4 g_j^2)} - \frac{f_i \delta_i(j)}{2 N g_j^2} - \frac{f_i^2}{4 N^2 g_j^2} \right) u_{js} + \left( \frac{\delta_i(j)}{g_j} - \frac{f_i}{N} \right) u_{js}^{(1)} + u_{js}^{(2)}.
\end{align*}
\]  
(3.16)

Using these differential coefficients we can evaluate the influence of each observation on the eigenvalues and the corresponding scores assigned to the categories. That is, when the weight for the individual \(i\) is slightly changed from 1 to 1-\(\varepsilon\), the eigenvalues and the scores are changed as follows.
\[
\theta_s \rightarrow \theta_s + \varepsilon \theta_s^{(1)} + \varepsilon^2 \theta_s^{(2)} + \cdots
\]  
(3.20)
\[ y_{js} = y_{js} + \varepsilon y_{js}^{(1)} + \varepsilon^2 y_{js}^{(2)} + \cdots \] (3.21)

As summarized measures of the change of scores \( \{y_{js}\} \) we shall use the following two values to evaluate the amount of influence on the scores \( \{y_{js}\} \) assigned to the categories.

(1) Euclidean norm of the difference of the two matrices \( Y_1 = (y_{is}) \) and \( Y_2 = (y_{js} + \varepsilon y_{js}^{(1)} + \varepsilon^2 y_{js}^{(2)} + \cdots) \):

\[
\| \varepsilon y_{js}^{(1)} + \varepsilon^2 y_{js}^{(2)} + \cdots \| = \left( \sum_{j=1}^{q} \sum_{s=1}^{q} (\varepsilon y_{js}^{(1)} + \varepsilon^2 y_{js}^{(2)} + \cdots)^2 \right)^{1/2},
\]

where \( q \) is the number of dimensions of interest.

(2) \( RV \) coefficient between the same two matrices

\[
RV(Y_1, Y_2) = \frac{\| \hat{Y}_1 \cdot \hat{Y}_2 \|^2}{\| \hat{Y}_1 \| \cdot \| \hat{Y}_2 \|},
\]

where \( \hat{Y}_1 \) and \( \hat{Y}_2 \) are the matrices made from \( Y_1 \) and \( Y_2 \) by centering column-wisely.

The \( RV \) coefficient is a measure of closeness between two configurations, which is invariant under any shift or rotation of axes for one or both configurations (Robert and Escoufier, 1976).

The changed values of scores \( \{x_{is}\} \) are calculated from the changed values of eigenvalues \( \{\lambda_s\} \) and scores \( \{y_{js}\} \) by using the equation (2.9).

<table>
<thead>
<tr>
<th>Category</th>
<th>Individual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1  2  3  4  5  6  7  8  9  10</td>
</tr>
<tr>
<td>1</td>
<td>1  1  1  0  0  1  1  1  1  0</td>
</tr>
<tr>
<td>2</td>
<td>0  0  0  0  1  1  0  0  0  1</td>
</tr>
<tr>
<td>3</td>
<td>1  1  1  1  0  1  1  1  1  0</td>
</tr>
<tr>
<td>4</td>
<td>0  1  1  1  1  1  1  1  1  1</td>
</tr>
<tr>
<td>5</td>
<td>1  1  1  1  1  1  1  1  1  1</td>
</tr>
<tr>
<td>6</td>
<td>1  1  0  0  0  1  1  0  0  0</td>
</tr>
<tr>
<td>7</td>
<td>1  1  1  1  0  1  1  0  1  0</td>
</tr>
<tr>
<td>8</td>
<td>0  1  1  1  1  0  1  1  1  1</td>
</tr>
<tr>
<td>9</td>
<td>1  1  1  1  1  0  0  0  0  1</td>
</tr>
<tr>
<td>10</td>
<td>1  1  0  1  1  1  0  1  1  1</td>
</tr>
<tr>
<td>11</td>
<td>0  0  1  1  0  0  1  1  1  1</td>
</tr>
<tr>
<td>12</td>
<td>0  1  1  1  1  0  1  1  1  1</td>
</tr>
</tbody>
</table>

Notes
(3) The figure "1" indicates that the food is accepted in the group in the sense that the mean scores using a 9-point scale is larger than 6.0 and the figure "0" indicates the opposite.
3.2 Evaluation of the influences of multiple observations

The methodology used in the case of a single observation can also be applied to the case of multiple observations.

We shall consider a set of individuals \( \mathcal{I} \) and introduce an indicator variable such that

\[
\phi_i = \begin{cases} 
1 & \text{if } i \in \mathcal{I}, \\
0 & \text{if } i \notin \mathcal{I}.
\end{cases}
\]  \hspace{1cm} (3.24)

We shall change the weight for the individuals \( \{i \mid i \notin \mathcal{I}\} \) from 1 to \( 1 - \varepsilon \). Then, like in the case of a single observation we obtain an expansion of the matrix \( H(\varepsilon) = (h_{ij}(\varepsilon)) \),

\[
h_{ij}(\varepsilon) = h_{ij} + \varepsilon h_{ij}^{(1)} + \varepsilon^2 h_{ij}^{(2)} + O(\varepsilon^3),
\]  \hspace{1cm} (3.25)

where

\[
h_{ij}^{(1)} = -\sum_{j=1}^{n} \phi_i \delta_j(\varepsilon) \delta_i(\varepsilon) + \frac{h_{ij}}{2} \sum_{j=1}^{n} \phi_i \left( \frac{\delta_j(\varepsilon)}{g_j} + \frac{\delta_i(\varepsilon)}{g_i} \right),
\]  \hspace{1cm} (3.26)

<table>
<thead>
<tr>
<th>Axis(s)</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eigenvalue ( (\lambda) )</td>
<td>0.176</td>
<td>0.132</td>
</tr>
<tr>
<td>Scores assigned to the categories ( (y_{ij}) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.598</td>
<td>-1.102</td>
<td></td>
</tr>
<tr>
<td>0.626</td>
<td>-0.299</td>
<td></td>
</tr>
<tr>
<td>0.027</td>
<td>0.990</td>
<td></td>
</tr>
<tr>
<td>-0.399</td>
<td>0.692</td>
<td></td>
</tr>
<tr>
<td>-1.843</td>
<td>-1.283</td>
<td></td>
</tr>
<tr>
<td>0.791</td>
<td>-1.817</td>
<td></td>
</tr>
<tr>
<td>0.818</td>
<td>0.664</td>
<td></td>
</tr>
<tr>
<td>-0.221</td>
<td>1.069</td>
<td></td>
</tr>
<tr>
<td>0.127</td>
<td>0.961</td>
<td></td>
</tr>
<tr>
<td>-1.782</td>
<td>-0.512</td>
<td></td>
</tr>
<tr>
<td>Scores assigned to the individuals ( (x_{ij}) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.282</td>
<td>0.183</td>
<td></td>
</tr>
<tr>
<td>-2.251</td>
<td>-3.317</td>
<td></td>
</tr>
<tr>
<td>1.003</td>
<td>0.399</td>
<td></td>
</tr>
<tr>
<td>-0.492</td>
<td>0.142</td>
<td></td>
</tr>
<tr>
<td>-0.062</td>
<td>-0.175</td>
<td></td>
</tr>
<tr>
<td>2.284</td>
<td>-1.758</td>
<td></td>
</tr>
<tr>
<td>1.221</td>
<td>0.035</td>
<td></td>
</tr>
<tr>
<td>-0.789</td>
<td>0.786</td>
<td></td>
</tr>
<tr>
<td>-0.704</td>
<td>-0.695</td>
<td></td>
</tr>
<tr>
<td>-0.329</td>
<td>-0.789</td>
<td></td>
</tr>
<tr>
<td>-0.568</td>
<td>1.774</td>
<td></td>
</tr>
<tr>
<td>-0.789</td>
<td>0.786</td>
<td></td>
</tr>
</tbody>
</table>

**Note** The eigenvalues obtained are \( \lambda_1 = 0.176, \lambda_2 = 0.1519, \lambda_3 = 0.0567, \lambda_4 = 0.0262, \ldots \). Since the values \( \lambda_1, \lambda_2, \ldots \) are relatively smaller than \( \lambda_1 \) and \( \lambda_2 \), we use two dimensional scores \((y_{11}, y_{12})\) and \((x_{11}, x_{12})\).
Fig. 1 Change of the configuration of 10 categories (groups) \( \{y_{js}\} \rightarrow \{y_{js} + \varepsilon y_{js}'\}, \) \( (\varepsilon = 0.3) \), when the weight for individual 2 is slightly changed from 1 to \( 1 - \varepsilon \) (Example 1)

Fig. 2 Change of the configuration of 12 individuals (foods) \( \{x_{is}\} \rightarrow \{x_{is} + \varepsilon x_{is}'\}, \) \( (\varepsilon = 0.3) \), when the weight for individual 2 is slightly changed from 1 to \( 1 - \varepsilon \) (Example 1)
Thus, by substituting (3.26) and (3.27) into (3.14) and (3.15) and by modifying (3.18) and (3.19) as

\[
y^{(1)}_{ji} = \sqrt{\frac{N}{g_j}} \left[ \left( \frac{3}{4g_j} \sum_i \phi_i \delta_i(j) - \frac{1}{2N} \sum_i \phi_i f_i \right) u_{js} + u^{(1)}_{js} \right],
\]

\[
y^{(2)}_{ji} = \sqrt{\frac{N}{g_j}} \left[ \frac{3}{4g_j} \left( \sum_i \phi_i \delta_i(j)^2 \right) - \frac{1}{2Ng_j} \left( \sum_i \phi_i \delta_i(j) \right) \left( \sum_i \phi_i f_i \right) \right]
- \frac{1}{4Ng_j} \left( \sum_i \phi_i f_i \right)^2 u_{js} + \left( \frac{1}{g_j} \sum_i \phi_i \delta_i(j) - \frac{1}{N} \sum_i \phi_i f_i \right) u^{(1)}_{js} + u^{(2)}_{js},
\]

we can evaluate the influences on the eigenvalues and the scores assigned to categories.

We can easily verify that \( h^{(1)}_{ji} \) in (3.24) can be obtained by adding \( h^{(1)}_{ji} \) in (3.7) for the individuals included in \( \mathcal{S} \). So if we consider only the first order term, we can easily calculate the results for multiple individuals from those for a single individual.

### 4. Numerical examples

#### Example 1.
Table 2 is a part of the result of the survey on food acceptance (Toda and Tanaka, 1968). It shows whether each of 12 foods is accepted or not in each of 10 groups, which are constructed by age and sex.

First let us apply Hayashi’s third method of quantification to these data. The common eigenvalues \( \lambda^2 \) of (2.6) and (2.7) and the corresponding quantities \{\( y_{js} \)\} and \( \{\lambda_{js} \} \) are

| Table 4 |
|-----------------|-----------------|-----------------|-----------------|
| The amount of influence of each food (Example 1) |
| \( \theta^{(1)}_i \) | \( \theta^{(2)}_i \) | \( \|y^{(1)}_{js}\| \) | \( RV \) |
| 1 | 0.01509 | 0.01233 | 2.959 | 0.9992 |
| 2 | -0.01677 | -0.04362 | 11.398 | 0.9961 |
| 3 | -0.00750 | 0.01146 | 3.761 | 0.9996 |
| 4 | 0.01233 | 0.01426 | 1.659 | 0.9994 |
| 5 | 0.2307 | 0.01597 | 1.176 | 1.0000 |
| 6 | -0.0485 | -0.01149 | 7.483 | 0.9983 |
| 7 | 0.01263 | 0.01165 | 2.315 | 0.9992 |
| 8 | 0.00604 | 0.00161 | 4.764 | 0.9994 |
| 9 | 0.01406 | 0.00278 | 2.568 | 0.9974 |
| 10 | 0.01993 | 0.00641 | 2.125 | 0.9989 |
| 11 | 0.00447 | -0.02299 | 4.260 | 0.9868 |
| 12 | 0.00604 | 0.00161 | 4.764 | 0.9994 |

**Notes**

1. The Euclidean norm \( \|y^{(1)}_{js}\| \) is calculated as

\[
\|y^{(1)}_{js}\| = \left( \sum_i \sum_j (y^{(1)}_{js})^2 \right)^{1/2}.
\]

2. The RV -coefficients \( RV (y_{js}, \lambda_{js} + \epsilon y^{(1)}_{js}) \) are calculated by setting \( \epsilon = 0.3 \).
Fig. 3 Change of the configuration of 10 categories (groups) \( \{y_{x_k}\} \rightarrow \{y_{x_k} + \varepsilon y_{x_k}^{(1)}\} \), \( \varepsilon = 0.3 \), when the weights for individuals 2 and 6 are slightly changed from 1 to \( 1 - \varepsilon \) (Example 1)

Table 5

Food acceptance data—acceptance patterns of 12 drinks in 10 groups (Example 2)

<table>
<thead>
<tr>
<th>Category Individual</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Notes

(1) Individuals (drinks): 1. wine, 2. beer, 3. green tea, 4. black tea, 5. coffee, 6. milk, 7. Calpis (lactic acid drink), 8. orange juice, 9. powdered juice, 10. coke, 11. soda pop, 12. Nectar (fruit juice)

Fig. 4 Change of the configuration of 10 categories (groups) \( \{y_{js}\} \rightarrow \{y_{js} + \varepsilon y_{js}'\} \), (\( \varepsilon = 0.3 \)), when the weight for individual 2 is slightly changed from 1 to \( 1 - \varepsilon \) (Example 2)

Fig. 5 Change of the configuration of 10 categories (groups) \( \{y_{js}\} \rightarrow \{y_{js} + \varepsilon y_{js}'\} \), (\( \varepsilon = 0.3 \)), when the weight for individual 2 is slightly changed from 1 to \( 1 - \varepsilon \) (Example 2)
Observing that the values $A_3$, $X_14$, ... are relatively smaller than $A_1$, $A_2$, we neglect the third and higher dimensions and use two-dimensional scores $\{y_1, y_2\}$ and $\{x_1, x_2\}$. The configurations of the categories $\{y_1, y_2\}$ and the individuals $\{x_1, x_2\}$ are shown as small circles in Fig. 1 and 2, respectively. Looking at the configuration of the categories, we may interpret that the first axis reflects young vs. old and the second axis reflects the middle age vs. the others. From the configuration of the individuals (foods) we may say:

1. "Croquette" and "iced noodles" are located far from the others.
2. "Croquette", "curry and rice", "ham" and "fried noodles" are located right. So they suit to the taste of younger people.
3. "Iced noodles", "sliced raw fish" and "broiled eel" are located left. So they suit to the taste of older people.

Next let us apply the sensitivity analysis developed in the preceding section.

In order to evaluate the influence of each individual on the configuration of the categories we calculate the differential coefficients $\{\theta_{11}, \theta_{12}\}$ and $\{y_1, y_2\}$, and summarize these into the four values shown in Table 4. According to this table it seems that the individuals (foods) 2 and 6 are influential. Fig. 1 shows the change of the configuration of the categories $\{y_1, y_2\} \rightarrow \{y_1 + \varepsilon y'_1\}, \varepsilon = 0.3$, due to a small change of the weight for individual 2. The corresponding change of the configuration of the individuals $\{x_1, x_2\} \rightarrow \{x_1 + \varepsilon x'_1\}, \varepsilon = 0.3$, is given in Fig. 2.

In these figures there occur little rotations of the configurations but the changes are so small that they do not affect the interpretations. Fig. 3 shows the change of the configuration of the categories $\{y_1, y_2\} \rightarrow \{y_1 + \varepsilon y'_1\}, \varepsilon = 0.3$, due to a small change of the weights for multiple individual No. 2 and 6. Also in this figure the change is not large.

Example 2. Table 5 is another part of the result of the same survey on food acceptance. Since the eigenvalues are obtained as

$$\lambda_1^2 = 0.292 > \lambda_2^2 = 0.184 > \lambda_3^2 = 0.165 > \lambda_4^2 = 0.081 > \lambda_5^2 = 0.038 > \cdots,$$

we neglect the fourth and higher dimensions and use three dimensional scores.

Table 6

<table>
<thead>
<tr>
<th>$\theta_{11}$</th>
<th>$\theta_{12}$</th>
<th>$\theta_{13}$</th>
<th>$|y'_1|$</th>
<th>$RV$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.0996</td>
<td>-0.03325</td>
<td>-0.0296</td>
<td>8.972</td>
</tr>
<tr>
<td>2</td>
<td>-0.05806</td>
<td>-0.0223</td>
<td>-0.0281</td>
<td>20.853</td>
</tr>
<tr>
<td>3</td>
<td>0.04172</td>
<td>0.01757</td>
<td>0.0381</td>
<td>10.925</td>
</tr>
<tr>
<td>4</td>
<td>0.0890</td>
<td>-0.01356</td>
<td>-0.01281</td>
<td>14.006</td>
</tr>
<tr>
<td>5</td>
<td>-0.01397</td>
<td>0.02274</td>
<td>-0.0162</td>
<td>5.309</td>
</tr>
<tr>
<td>6</td>
<td>0.04523</td>
<td>0.01589</td>
<td>0.02184</td>
<td>3.849</td>
</tr>
<tr>
<td>7</td>
<td>0.0366</td>
<td>0.01246</td>
<td>0.00849</td>
<td>7.733</td>
</tr>
<tr>
<td>8</td>
<td>0.02403</td>
<td>0.00181</td>
<td>0.01162</td>
<td>5.042</td>
</tr>
<tr>
<td>9</td>
<td>-0.0423</td>
<td>0.00682</td>
<td>-0.01399</td>
<td>3.680</td>
</tr>
<tr>
<td>10</td>
<td>0.00943</td>
<td>0.02382</td>
<td>-0.00624</td>
<td>11.873</td>
</tr>
<tr>
<td>11</td>
<td>-0.02384</td>
<td>0.00862</td>
<td>0.00939</td>
<td>2.682</td>
</tr>
<tr>
<td>12</td>
<td>-0.0423</td>
<td>0.00682</td>
<td>-0.00662</td>
<td>3.680</td>
</tr>
</tbody>
</table>
The influence of each individual on the eigenvalues and scores are summarized in Table 6. The locations of the categories and their movements due to a small change of the weight for the most influential individual (No. 2 = beer) are shown in Fig. 4 and 5. The locations of ten categories especially on the 1-2 plane suggest us the existence of three clusters such as the children cluster, the adult men cluster and the adult women cluster. However, after the weight for individual 2 is changed, the distinction of the three clusters is not so clear on the 1-2 plane. It may be due to the rotation on the 2-3 plane appeared in Fig. 5.

5. Discussion

We have investigated how the result of Hayashi's third method of quantification changes when the weight for a single or multiple individuals is changed slightly.

We considered up to the first order term of \( \varepsilon \) and used the differential coefficients \( \{ \theta_0^{(i)} \}, \{ y_0^{(i)} \} \) and \( \{ x_0^{(i)} \} \) to evaluate the influences on the eigenvalues and the corresponding scores in the above examples. If the sample size is sufficiently large so that \( 1/f_i < 1 \) and

\[
\frac{\delta_r(j) \delta_r(j')}{f_r} / \sum_{i=1}^n \frac{\delta_r(j) \delta_r(j')}{f_r} \ll 1 \]

for any \( j \) and \( j' \), the values \( \theta_0 + \theta_0^{(i)} \) and \( y_0 + y_0^{(i)} \) give the approximations to the results when the individual \( i \) is deleted. So we may consider that Fig. 1, 4 and 5 indicate approximately 3/10 times the changes due to the deletion of the individual \( i \). We can improve these approximations by using the second order terms in the power series expansions of the eigenvalues and the corresponding scores given in section 3. To examine the effect of the second order term we show the result of sensitivity analysis of the first example obtained by using up to \( \varepsilon^2 \) in Fig. 6. Comparing Fig. 6 with Fig. 1 we may say that the effect of the second order term is not so large even in our
example of small sample.

As measures for the amount of influence we used (1) the differential coefficient \( \theta_{ij}^{(1)} \) of the \( s \)-th eigenvalue, (2) the Euclidean norm \( \| y_{js}^{(1)} \| \) of the differential coefficients of \( y_{js} \)-scores, and (3) the RV-coefficient between the two configurations \( \{ y_{js} \} \) and \( \{ y_{js} + \varepsilon y_{js}^{(1)} \} \). The first measure indicates the change of the proportion explained in the \( s \)-th dimension. The other two indicate the change of the configuration \( \{ y_{js} \} \). The Euclidean norm \( \| y_{js}^{(1)} \| \) is simply the total length between \( \{ y_{js} \} \) and \( \{ y_{js} + y_{js}^{(1)} \} \), while the RV-coefficient is a measure which is invariant under any orthogonal transformation of \( \{ y_{js} + \varepsilon y_{js}^{(1)} \} \) and becomes zero when an orthogonal transform of \( \{ y_{js} + \varepsilon y_{js}^{(1)} \} \) coincides with \( \{ y_{js} \} \). So when a rotation occurs, the change measured by the RV-coefficient is small, though the change measured by the Euclidean norm in large. Such phenomena were observed in our two examples, especially in Example 2.

For investigation of the influence of multiple individuals it is not realistic to consider all possible combinations of individuals because the number of combinations is extraordinarily large. As is shows in section 3.2 the first derivatives of a set of individuals is evaluated by adding the derivatives of each individual belonging to the set. Hence it is expected that the influence of a set of individuals which have similar \( \{ y_{js}^{(1)} \} \) values becomes large. To collect the similar individuals it may be useful to apply the principal component analysis or the cluster analysis to the \( n \times Rq \) data matrix \( G \) of the first derivatives.

In Hayashi's third method of quantification the formulation is symmetric for rows (individuals) and columns (categories). Therefore we can evaluate the influence of each category by changing the position of individuals and categories in the preceding formulation.

**REFERENCES**


(Received March, 1984)