CORRESPONDENCE ANALYSIS OF AN ARTIFICIAL TORUS DATA

Masashi Okamoto*

As artificial data sets for correspondence analysis, a torus data and a solid torus data are presented. The former has two circular traits and the latter two circular traits and one linear trait representing the radius. The eigenvalue problem for each data is shown to be solvable analytically by dealing with the two parameters each describing the circular trait as free parameters. The solution shows a competition of eigenvalues of the traits involved, so-called Guttman effect.

1. Introduction

An artificial data might be useful in investigating methodolgical properties of any statistical procedure. As an artificial data for correspondence analysis, Iwatsubo (1984) presented a circular data. The author developed his idea by proposing a disk data in Okamoto (1994b) and a cylinder data as well as a solid cylinder data in Okamoto (1994c). Each of the author’s first two data sets has one linear trait and one circular trait, while the last one has two linear traits and one circular trait.

This paper first deals with a torus data which arises when two circles located at the both ends of a cylinder are identified. The data has two circular traits. The paper then treats a solid torus data which is generated similarly from a solid cylinder data and has two circular traits and one linear trait which represents the radius. The eigenvalue problem for the former data can be solved analytically quite easily. The latter is analyzable whenever the maximum radius is a small positive integer.

Iwatsubo (1984) showed that, similarly as for a linear trait, a circular trait has the Guttman series; intensity, closure, involution, etc. besides the main effect. Okamoto (1992, 1994b and 1994c) showed that, in correspondence analysis of a data involving more than one trait, a severe competition between traits may take place in terms of the Guttman series. To be specific, a certain trait may be so strong that some member of its Guttman series precedes some other trait(s) in the decreasing sequence of eigenvalues. When a statistician tries to interpret the result of his/her analysis, a certain number of axes starting from the largest eigenvalue are usually adopted. Such a competition, however, implies a pitfall in that some important

Key Words and Phrases: Artificial data, circular trait, correspondence analysis, eigenvalue problem, Guttman effect, solid torus data, torus data.

trait(s) contained in the data might be missed. It is the case also with torus data and solid torus data.

2. Artificial torus data

Assume that a geometric figure consisting of points and lines is given, where every line connects two points. Let us regard a point as an item and a line as a subject and suppose that a subject responds to an item if the associated line has the associated point as one of the two end-points. Then a data matrix of the type of subject \( \times \) item emerges and it can be analyzed by correspondence analysis (see e.g. Greenacre, 1984). The original figure will be called item pattern. The data is obtained with no observation error and the problem is whether a scatter plot derived by the analysis reproduces the item pattern faithfully.

Denote the number of items by \( m \). Suppose for an \( m \times m \) matrix

\[
D = (d_{ij}),
\]

the diagonal element \( d_{ij} \) denotes the number of responses for the item \( j \) and the nondiagonal element \( d_{ij} (=1 \ or \ 0) \) shows whether or not there exists a subject which responds to both items \( i \) and \( j \). The matrix will be called the item response matrix. Let \( G \) be a diagonal matrix consisting of the diagonal elements of \( D \). Then the eigenvalue problem is defined as a matrix equation

\[
Db = \lambda Gb,
\]

where \( \lambda \) is an eigenvalue and \( b \) is an eigenvector. Scores of each item can be obtained by standardizing \( b \) so as to have unit variance.

Let us consider a cylinder data due to Okamoto (1994c) with \( k \) points on a circle and length \( q \), where each item can be described by

\[(j, p) \ for \ j=1, \ldots, k \ and \ p=1, \ldots, q+1\]

with the coordinate \( j \) indicating the position on a circle and \( p \) the number of the circle carrying the item, counted from one end of the cylinder on. Now if the two items with the coordinates \((j, 1)\) and \((j, q+1)\) located on the both ends of the cylinder are identified for every \( j \), then the items are expressed by

\[(j, p) \ for \ j=1, \ldots, k \ and \ p=1, \ldots, q,\]

(2.3)
giving rise to an item pattern of a torus data with the two parameters \( k \) and \( q \). Each parameter represents a circular trait. The trait represented by \( k \) and \( q \) will be denoted by C and T, respectively. The number of items becomes \( m = kq \), that of subjects is \( n = 2kq \) and the total number of responses is \( N = 4kq \). The item pattern with \( k = q = 4 \) is shown as Fig. 1.

When the items are arranged in a natural order, the matrix \( D \) in the eigenvalue problem (2.2) is written as
where symmetric matrices of order $k$ are arranged as blocks in both horizontal and perpendicular directions, $q$ each in number. The block $E=(e_{ij})$ on the diagonal is defined by

\[
E_{ii}=4 \quad \text{for} \quad i=1, \ldots, k, \\
E_{i,i+1}=E_{i+1,i}=1 \quad \text{for} \quad i=1, \ldots, k-1, \\
e_{1k}=e_{k1}=1, \\
e_{ij}=0 \quad \text{otherwise}.
\]

The identity matrix $I_k$ appears on the both adjacent places of the diagonal, at the right uppermost and the left lowermost corners. Any other block is occupied by a zero matrix.

The eigenvector $b$ is to be standardized in the mean and variance with respect to the matrix $G$. The condition for variance is written as

\[
b'Gb=N.
\]

Denote by $1_k$ a $k$-vector with unit components. The trivial solution $\lambda=1$ with $b=1_k$ not allowing standardization in the mean will be excluded so that there exist $(m-1)$ linearly independent solutions.
3. Eigenvalue analysis of a torus data

It seems strange but is actually natural that a circular trait is simpler than a linear trait analytically. The reason is that every point on a circle is equivalent with each other, while each point on a line has its own character. In Okamoto (1994c) it was proved difficult to solve the eigenvalue problem for a cylinder data analytically with \( L \) as a free parameter so that it was solved for some particular small values of \( L \), starting from 1. But a torus data will be seen to be solvable with both \( k \) and \( q \) as free parameters. First define

\[
C_t = \cos \left( \frac{2t\pi}{k} \right) \quad \text{for} \quad t = 0, 1, \ldots, \left[ \frac{k}{2} \right],
\]

\[
C_* = \cos \left( \frac{2z\pi}{q} \right) \quad \text{for} \quad z = 0, 1, \ldots, \left[ \frac{q}{2} \right],
\]

where \( \left[ \right] \) denotes the Gauss symbol. In the following theorems the score vector becomes always a direct product of two parts and the score corresponding to the item (2.3) is written as

\[
b_{\alpha} = xu_{\alpha}w_{\alpha} \quad \text{for} \quad j = 1, \ldots, k \quad \text{and} \quad p = 1, \ldots, q
\]

where \( x \) is a proportional factor determined by the condition (2.6).

**Theorem 1**

The general form of eigenvalues for a torus data is given by

\[
\lambda_{\alpha} = \frac{1}{4} \left( 2 + C_t + C_* \right)
\]

with two parameters \( t \) and \( z \) with the domains stated in (3.1), excluding the case \( t = z = 0 \). The eigenvalue is a fourfold root when \( t \neq 0 \) and \( z \neq 0 \), a double root either when \( t \neq 0 \) and \( z = 0 \) or when \( t = 0 \) and \( z \neq 0 \). Furthermore it is a double root either when \( k \) is an even number and \( t = k/2 \) or when \( q \) is even and \( z = q/2 \). Finally it is a single root when both of the last two pairs of conditions are satisfied.

Score vectors associated with these eigenvalues are expressed as

\[
b_{\alpha} = (b_{\alpha})_{\alpha} = (xu_{\alpha}w_{\alpha})
\]

after (3.2), where

\[
u_{\alpha} = \cos (\theta_{\alpha}) \quad \text{or} \quad \sin (\theta_{\alpha}),
\]

\[
w_{\alpha} = \cos (\theta_{\alpha}^*) \quad \text{or} \quad \sin (\theta_{\alpha}^*),
\]

\[
\theta_{\alpha} = \pm 2jt\pi/k + \beta_t, \quad \theta_{\alpha}^* = \pm 2pz\pi/q + \beta_z^*,
\]

\( \beta_t \) and \( \beta_z^* \) being arbitrary constants. It is noted that

\[
x = 2 \quad \text{when} \quad t \neq 0 \quad \text{and} \quad z \neq 0,
\]

\[
x = \sqrt{2} \quad \text{and} \quad u_{j0} = 1 \quad \text{when} \quad t = 0 \quad \text{and} \quad z \neq 0 \quad \text{for} \quad j = 1, \ldots, k,
\]

\[
x = \sqrt{2} \quad \text{and} \quad w_{p0} = 1 \quad \text{when} \quad t \neq 0 \quad \text{and} \quad z = 0 \quad \text{for} \quad p = 1, \ldots, q
\]

and that
\[ u_{jt} = \begin{cases} (-1)^t \text{ when } k \text{ is even and } t = k/2 \text{ for } j = 1, \ldots, k, \\ w_{pz} = (-1)^p \text{ when } q \text{ is even and } z = q/2 \text{ for } p = 1, \ldots, q \end{cases} \] 

and also that the factor \( x \) in (3.3) starts with 2 and is divided by \( \sqrt{2} \) in each case of (3.7).

The theorem exhausts all solutions because it provides linearly independent solutions, \( kq - 1 (= m - 1) \) in number. The vector \( b_{t0} \) represents the circular trait \( C_t \) since its components are independent of \( p \), where \( t = 1 \) indicates the main effect \( C = C_1 \), \( t = 2 \) the Guttman intensity \( C_2 \) of \( C \), \( t = 3 \) closure \( C_3 \) and then the remaining Guttman series of \( C \) after that. The vector \( b_{0z} \) represents the circular trait \( T_z \) since its components are independent of \( j \), where \( z = 1 \) indicates the main effect \( T = T_1 \), \( z = 2 \) the Guttman intensity \( T_2 \) of \( T \) and then the remaining Guttman series. The vector \( b_{tz} \) with \( t > 0 \) and \( z > 0 \) may be called interaction because its components given by (3.3) is a direct product of the scores of the traits \( C_t \) and \( T_z \). It will be denoted by \( (C_t T_z) \). Each of \( \beta_t \) and \( \beta_z \) shows freedom of rotation in the 2-dimensional spaces spanned by the two eigenvectors associated with the equal eigenvalues. Each of the two double signs \( \pm \) in (3.6) determines the direction of rotation.

Suppose \( \beta_t = 0 \) in (3.6). If \( \cos \) is adopted in (3.4), a formula in trigonometric functions implies that \( u_{t2} \) as a function of \( j \), is quadratic in \( u_{jt} \) which implies in turn that the scatter plot using the two axes becomes a quadratic curve. This phenomenon for a linear trait is called arch effect or horseshoe effect (Greenacre, 1984). But such a wording seems inadequate with a circular trait. It might better be called the Guttman effect.

The score vector when \( k \) is even and \( t = k/2 \) coincides with the result obtained by adopting \( \cos \) in (3.4), putting \( t = k/2 \) and \( \beta_t = 0 \) and then dividing the proportional factor by \( \sqrt{2} \). It is also the case with the trait \( T \) and the situation is common with all theorems in Okamoto (1994b and 1994c).

It is interesting that the set of the eigenvalues in Theorem 1 does not change under interchanging values of the two parameters \( k \) and \( q \) because connections between items do not change, though the 3-dimensional figure in Fig. 1 changes. In order to see the ordering of traits appearing in the sequence of eigenvalues, the result of identifying the first twenty axes is shown as Table 1 with the domain of \( k = 5 \) and 7 and \( q = 3 \) and 5 (\( k > q \)). The trait \( T \) is concealed by \( C_2 \), the intensity of

<table>
<thead>
<tr>
<th>( k )</th>
<th>( q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>CTC_2 (CT) (C_2T)</td>
</tr>
<tr>
<td>7</td>
<td>CC_2T (CT) C_1 (C_1T) (C_1T)</td>
</tr>
</tbody>
</table>
C, at the entry \((k, q) = (7, 3)\). Thus it is required to remove the Guttman effect \(C_2\) to get \(T\). It is also found that axes related to \(C\) are advanced (with reduced ranks) in the sequence when \(k\) is increased, while axes related to \(T\) are advanced when \(q\) is increased.

4. **Solid torus data**

A solid cylinder data by Okamoto (1994c) is obtained by adding the center to a cylinder data and by allowing to extend the item pattern to the direction of radiiuses. Let \(k\) be the number of points on a circle and \(q\) the length of the cylinder as before. Adopting the new parameter \(r\) representing the radius, each item is described by using the two coordinates

\[
(j, p) \text{ for } j = 0, 1, \ldots, kr \text{ and } p = 1, \ldots, q + 1.
\]

The center corresponds to \(j = 0\), the points on the smallest circles to \(j = 1, \ldots, k\) and then follow the points on larger circles. If the two items \((j, 1)\) and \((j, q + 1)\) located at the both ends of the solid cylinder are identified for every \(j\), then the items are described as

\[
(j, p) \text{ for } j = 0, 1, \ldots, kr \text{ and } p = 1, \ldots, q
\]

which give rise to the item pattern of a solid torus data with three parameters \(k\), \(q\) and \(r\). The torus data up to the last section has been so to speak void, while the new one is solid. Two parameters \(k\) and \(q\) denote circular trait \(C\) and \(T\), respectively, as before, while \(r\) denotes a linear trait \(R\) representing the radius. The number of items is \(m = (kr + 1)q\) and that of subjects is \(n = (3kr + 1)q\). The case when \(k = q = 4\) and \(r = 1\) is shown as Fig. 2.

If all items are arranged in a natural order, then by defining \(h = kr + 1\) the matrix \(D\) in the eigenvalue problem (2.2) is written as

![Fig. 2 Item pattern of a solid torus data \((k = q = 4\) and \(r = 1\)](image-url)
It is a symmetric block matrix of type $q \times q$ consisting of blocks which are symmetric matrices of order $h$. Similarly as in (2.4), symmetric matrices $E_h$ appear on the diagonal and identity matrices $I_h$ appear on the two adjacent places and at the right uppermost and left lowermost corners. Zero matrices appear elsewhere.

The form of the matrix $E_h$ is determined by the value of $r$. When $r = 1$

$$E_h = \begin{bmatrix} k+2 & 1_k' & 1_k \\ 0_k' & E+I_h \\ 1_k' & I_h \\ 0_k & I_h \end{bmatrix}$$

with $E$ defined in (2.5). If $r = 2$, then

$$E_h = \begin{bmatrix} k+2 & 1_k' & 0_k' \\ 1_k & E+2I_h & I_h \\ 0_k & I_h & E+I_h \end{bmatrix}$$

where $0_k$ is a $k$-vector consisting of components 0.

It is difficult for a solid torus data to get a general analytical solutions with $r$ as a free parameter. As in Okamoto (1994b and 1994c) let us deal with the problem for some specific small values of $r$, starting from 1.

**Theorem 2**

When $r = 1$ for a solid torus data, there are three types of eigenvalues. First, the type R is given by

$$\lambda_R = \frac{2}{5} + \frac{1}{k+2}.$$  

Eigenvalues of type C are given by

$$\lambda_C = \frac{1}{10} \left( 5 + 2C_t + 2C_z^* \right)$$

for $t = 1, \cdots, [k/2]$ and $z = 0, 1, \cdots, [q/2]$, where $C_t$ and $C_z^*$ are defined in (3.1). Each of them is a double root when $z = 0$ and a fourfold root when $z \neq 0$. Eigenvalues of type T are given by the two solutions $\lambda_T^{\pm}$ of the quadratic equation in $\lambda$

$$\left( (k+2)(1-2\lambda)+2C_z^* \right) \left( (7-10\lambda)+2C_z^* \right) = k$$

for $z = 1, \cdots, [q/2]$. Each solution is a double root.
Associated with these eigenvalues, the vector $\mathbf{b}_R$ of type R is a direct product of the form (3.2), determined by
\[
\begin{align*}
&u_0=5k \text{ and } u_1=\cdots=u_k=-(k+2), \\
&w_1=\cdots=w_k.
\end{align*}
\] (4.8)

The vector $\mathbf{b}_{Ct}$ of type C is a direct product of the form (3.3), where
\[
\begin{align*}
&u_{0i}=0 \text{ and } u_{jt} \text{ given in (3.4)}, \\
&w_{pz}=1 \text{ when } z=0 \text{ and (3.5) otherwise}.
\end{align*}
\] (4.9)

The vector of type T is a direct product of the form $\mathbf{b}_{Tiz}=(\mathbf{b}_{piz})=(xu_{iz}w_{pz})$, where
\[
\begin{align*}
&u_{0i}=\{(7-10\lambda_{Tiz})+2C_z^*\}, \\
&w_{pz} \text{ in (3.5)}
\end{align*}
\] (4.10)
for $i=1$ and 2. The factor $x$ can be determined by the condition (2.6) in each case.

The procedure when either $k$ or $q$ is an even number is the same as in Theorem 1.

The number of eigenvalues is one for type R, $(k-1)q$ for type C irrespective of whether $k$ or $q$ is even or odd and $2(q-1)$ for type T. Thus the total number is $(k+1)q-1$ (= $m-1$), exhausting all solutions.

Interpretation of score vectors is as follows. The vector $\mathbf{b}_R$ represents the trait R contrasting the center and the circumference because its components are independent of both $j \geq 1$ and $p$, $u_0$ and $u_i$ having different signs. The vector $\mathbf{b}_{Ct0}$ represents the circular trait C because its components are independent of $p$. As for the meaning of the parameter $t$, $t=1$ means the main effect $C_1$, $t=2$ intensity $C_2$ of C and the Guttman series $C_t$ in general. The vector $\mathbf{b}_{Tiz}$ represents the circular trait T because its components are independent of $j \geq 1$ and $u_{0i}>0$ after some calculation. As for the meaning of the parameter $z$, $z=1$ means the main effect $T_1$, $z=2$ intensity $T_2$ of T and the Guttman series $T_z$ in general. The vector $\mathbf{b}_{Tz}$ represents the interaction between $T_z$ and R because $u_{0i}<1$ so that it will be denoted by the symbol $(T_zR)$. The vector $\mathbf{b}_{Cz}$ for $z \neq 0$ represents the interaction (C$_t$T$_z$) between C$_t$ and T$_z$.

For each of the two groups $\{\lambda_{Ct}\}$ and $\{\lambda_{Tiz}\}$ the eigenvalue decreases as $t$, $z$ or $i$ increases. But if all eigenvalues including $\lambda_k$ are to be rearranged in the decreasing order of magnitude, the three groups are mixed in a complicated way. This issue will be discussed after Theorem 3.

Theorem 3

Suppose $r=2$ for a solid torus data. There are three types of eigenvalues. For the type R they are two solutions $\lambda_{R1}>\lambda_{R2}$ of the quadratic equation in $\lambda$
\[
60(k+2)\lambda^2-2(37k+104)\lambda+(20k+89)=0.
\] (4.11)

For the type C the eigenvalues are given by
\[
\lambda_{Cz} = \frac{1}{60} \{(30+11C_z \pm (30+C_z)\pm 1/2)\}.
\] (4.12)
for \( t = 1, \ldots, \lfloor k/2 \rfloor \) and \( z = 0, 1, \ldots, \lfloor q/2 \rfloor \), where \( C_{tz} = C_t + C_z^* \) based on (3.1). Each of them is a double root when \( z = 0 \) and a fourfold root when \( z \neq 0 \). The eigenvalues of type T are given by the three solutions \( \lambda_{1z} > \lambda_{2z} > \lambda_{3z} \) of the cubic equation

\[
\frac{(k+2)(1-2 \lambda) + 2C_z^*}{1 - 12 \lambda + 2C_z^*} \cdot \frac{k}{7-10 \lambda + 2C_z^*} = 0
\]

for \( z = 1, \ldots, \lfloor q/2 \rfloor \). All of them are double roots.

Associated with these eigenvalues, the eigenvector \( b_{Ri} \) \((i = 1 \text{ and } 2)\) of type R is a direct product of the form (3.2), where

\[
\begin{align*}
  u_1 & = \cdots = u_k \\
  u_{k+1} & = \cdots = u_{2k} (= u'), \\
  w_1 & = \cdots = w_q
\end{align*}
\]

and where the proportional relationship between \( u_0, u_1 \) and \( u' \) is determined by putting \( y_0 = u_0/u' \) and \( y_1 = u_1/u' \) and

\[
y_0 = (10 \lambda_{8i} - 9)(12 \lambda_{8i} - 10) - 1 \quad \text{and} \quad y_1 = 10 \lambda_{8i} - 9.
\]

The vector \( b_{Ct} \) of type C is a product of the form (3.3), where

\[
\begin{align*}
  u_{0t} & = 0, \quad u_{jt} \text{ in (3.4)} \\
  u_{j,t} & = y'u_{jt} \quad \text{for } j = 1, \ldots, k, \\
  w_{pz} & = 1 \text{ when } z = 0 \text{ and (3.5) otherwise}
\end{align*}
\]

with the proportional factor \( y' \) given by

\[
y' = 12\lambda_{Ct} - (6 + 2C_z^*).
\]

Finally the eigenvector of type T is given by \( b_{Tk} = (xu_jw_{iz}) \), where

\[
\begin{align*}
  u_1 & = \cdots = u_k \\
  u_{k+1} & = \cdots = u_{2k} (= u'), \\
  w_{iz} & = \text{ in (3.5)}
\end{align*}
\]

with the proportional relationship determined by

\[
\begin{align*}
  u_1 & = (10 \lambda_{Tk} - 7 - 2C_z^*)u', \\
  u_0 & = (12 \lambda_{Tk} - 8 - 2C_z^*)u_1 - u'.
\end{align*}
\]

The factor x is always determined by (2.6). The procedure when either \( k \) or \( q \) is even is the same as in Theorem 1.

The number of eigenvalues is two for type R, \( 2(k-1)q \) for type C irrespective of whether \( k \) or \( q \) is even or odd and finally \( 3(q-1) \) for type T. The total number is \( 2k + 1)(q-1) = (m-1) \), which exhausts all solutions.

The vector \( b_{R1} \) represents the main effect R because \( y_0 < 0 \) after some calculation when \( i = 1 \) in (4.15), while \( b_{R2} \) represents the Guttman intensity \( R_2 \) of R because \( y_0 > 0 \) and \( y_1 < 0 \) when \( i = 2 \). The vector \( b_{Ct0} \) represents the circular trait C, where \( t = 1 \) means the main effect \( C_1 \), \( t = 2 \) intensity \( C_2 \) and then the remaining Guttman series. The vector \( b_{Tkz} \) represents the circular trait T, where \( z = 1 \) means the main effect \( T_1 \), \( z = 2 \) intensity \( T_2 \) and then the remaining Guttman series. The vector \( b_{Ct} \)
when \( z \neq 0 \) represents the interaction \((C_tT_z)\) between \( C_t \) and \( T_z \). Finally the vector \( b_{TRZ} \) represents the interaction \((T_zR)\), while \( b_{TZR} \) the interaction \((T_zR)\).

Within each of the three groups of eigenvalues \( \{ \lambda_2 \}, \{ \lambda_{C12} \} \) and \( \{ \lambda_{T2} \} \), the eigenvalue is monotone decreasing in each of the parameters \( t, z \) and \( i \) but when all values are arranged in the decreasing order of magnitude, the three groups are intermingled in more complicated way than with Theorem 2.

Table 2 shows the ordering of traits in the sequence of axes up to the eighteenth when \( r = 2 \) and each of \( k \) and \( q \) takes two values 3 and 5. Three traits \( C, T \) and \( R \) appear at the first three axes except for the unique case whe \( k = 3 \) and \( q = 5 \), in which the circular trait \( C \) is concealed after the interaction \((T_zR)\) of \( T \) and \( R \). In this case \((T_zR)\) should be removed in order to get \( C \). It is seen that if \( k \) (or \( q \)) increases, the axes related to \( C \) (or \( T \)) are advanced in the sequence. It is also seen for the two entries when \( k = q \) in Table 2 that between the two circular traits \( C \) and \( T \) the former is weaker than the latter perhaps because it is related to the linear trait \( R \).

5. Proofs of the theorems

The following lemma can be proved by using the addition theorem for trigonometric functions.

Lemma

For \( u_{j,t} \) and \( w_{pz} \) satisfying (3.4) and (3.5), respectively, the following equalities hold for any values of \( j, p, t \) and \( z \) :

\[
\begin{align*}
   u_{j+1,t} + u_{j-1,t} &= 2C_t u_{j,t}, \\
   w_{p+1,z} + w_{p-1,z} &= 2C_z^* w_{pz},
\end{align*}
\]

where \( C_t \) and \( C_z^* \) are defined in (3.1).

Proof of Theorem 1

The element of the matrix \( D \) in (2.4) which corresponds to the pair of the item \((j, p)\) and the second item \((j', p')\) in the expression (2.3) is given by
\[ d_{jp'p''} = \begin{cases} e_{jp'} & \text{when } p' = p, \\ \delta_{jp'} & \text{when } p' = p \pm 1, \\ 0 & \text{otherwise,} \end{cases} \]  

(5.1)

where \( \delta \) denotes the Kronecker delta and the values 0 and \( q+1 \) of \( p' \) should be interpreted as \( q \) and 1, respectively, with modulus \( q \). Then (2.2) is written as

\[ \sum_{jp'} d_{jp'p''} b_{jp''} = 8 \lambda b_{jp} \text{ for every } j \text{ and } p, \]

(5.2)

where the summation is extended over the domain \( j' = 1, \cdots, k \) and \( p' = 1, \cdots, q \). It is sufficient to prove that \( b_{j'p'} = b_{jp} \) defined in (3.3) satisfies this relation. Using (5.1), the left-hand side of (5.2) can be written as

\[ (4 \text{ut} + \text{ut},+1,t + \text{ut}+1,t)wpz + \text{ut}(\text{wp},-1,z + \text{wp},-1,z) \]

up to the multiplicative factor \( x \). By using Lemma it can be written as

\[ (4 + 2C_1 + 2C_2^*) \text{ut}wpz, \]

which implies (5.2) for \( \lambda = \lambda_{1z} \). [QED]

Proof of Theorem 2

The expression (5.1) still holds but \( e_{jp'} \) are now the elements of the matrix \( E_1 \) defined in (4.3). Equation (2.2) then becomes

\[ \sum_{jp'} d_{jp'p''} b_{jp''} = 2\lambda d_{jp} b_{jp} \text{ for every } j \text{ and } p. \]

(5.3)

Let us first find a score vector \( b_1 \) of type R. Changing the domain of \( j \) in (3.2) into \( j = 0, 1, \cdots, k \), we assume

\[ u_1 = \cdots = u_k \text{ and } w_1 = \cdots = w_q. \]

(5.4)

As for \( j \) in (5.3) we have only to consider the two values 0 and 1 because of symmetry. Then the eigenvalue problem of dimension \( kq \) is reduced to the following matrix equation of dimension 2:

\[
\begin{bmatrix}
  k+4 & k \\
  1 & 9
\end{bmatrix}
\begin{bmatrix}
  u_0 \\
  u_1
\end{bmatrix}
= 2\lambda
\begin{bmatrix}
  k+2 & 0 \\
  0 & 5
\end{bmatrix}
\begin{bmatrix}
  u_0 \\
  u_1
\end{bmatrix}
\]

Using \( \lambda \neq 1 \), the simultaneous equation leads to the solution \( \lambda_r \) in (4.5). Substituting it into one of the equations, we have \((k + 2)u_0 + 5ku_1 = 0\), which implies (4.8).

For the type C we assume (4.9). Then (5.3) leads to (4.6) because of Lemma.

For the type T we assume

\[ u_1 = \cdots = u_k = 1 \text{ and } w_p \text{ in (3.5)}. \]

Then, using the two cases \( j = 0 \) and 1, (5.3) is reduced to

\[
\begin{bmatrix}
  k+2+2C_2^* & k \\
  1 & 7+2C_2^*
\end{bmatrix}
\begin{bmatrix}
  u_0 \\
  u_1
\end{bmatrix}
= 2\lambda
\begin{bmatrix}
  k+2 & 0 \\
  0 & 5
\end{bmatrix}
\begin{bmatrix}
  u_0 \\
  u_1
\end{bmatrix}
\]


This implies the quadratic equation (4.7) and (4.10). [QED]

Theorem 3 can be proved similarly but omitted because of space limitation.

6. Concluding remarks

This paper should not be regarded as an end as it is. It might be better regarded as only a beginning of the end. It was shown in Okamoto (1992) that in some data with a two-dimensional linear structure the appearance of a weaker linear trait is hindered by the Guttman intensity of the stronger linear trait. In order to challenge the difficulty, the author presented in Okamoto (1994a) a method to remove the Guttman intensity, closure etc. of such a stronger linear trait to rescue the main effect of a weaker trait.

Similar situation may occur for a circular trait as is shown in this paper and also in Okamoto (1994b and 1994c). If the data is of the type of disk, cylinder or torus exactly, then the task of removing the Guttman series of a circular trait is not difficult because each member of the series gives rise to pairs of identical eigenvalues. In practice, however, the circular trait will be blurred by irregular responses of subjects to items so that two eigenvalues cease to be identical, rendering the task complicated. Thus, it is an open problem at present.

Furthermore, there is a possibility that typical traits other than a linear or a circular one exist. Under this uncertainty it seems to be impossible to remove the Guttman series of all traits to secure the main effects of important traits for a given data set. In the author's opinion, correspondence analysis of any practical data requires prudent insight of the analyst in judging which axis among the set of eigenvectors is meaningful or not, undisturbed by the order of magnitude of eigenvalues.

There remains another important issue related to the Guttman effect. It is how to define the contribution ratio of each axis in correspondence analysis. Usually it is defined as the ratio of the corresponding eigenvalue divided by the sum of all eigenvalues in accordance with the custom in principal component analysis. But this policy seems inadequate for a data with a simple structure. Suppose that the data has only one trait, linear or circular. If it is linear, then the first axis contains all information involved in the data so that its contribution should be defined one. Similarly, if the trait is circular, the sum of the first two contribution ratios should be one. In both cases, however, the usual definition does not satisfy this requirement, yielding unreasonably small value as the contribution ratio for the main effect. The issue should be clarified.

The author thanks the referees for helpful comments.
REFERENCES


(Received November, 1993. Revised September, 1994)