LATENT SCALOGRAM ANALYSIS

Nobuoki Eshima*

In the present paper, scalogram analysis proposed by Guttman (1950) is developed into latent scalogram analysis, and a general discussion of the analysis is presented. The present paper deals with not only linear hierarchical structures but also branching hierarchical structures. In order to select a hierarchical structure which best fits the data set, model selection procedures are considered in both exploratory and confirmatory contexts. Concerning latent scales which exist in the population under consideration, a dynamic interpretation of latent scales is discussed through a mathematical viewpoint, and a method for evaluating the proportions of latent scales is proposed. Moreover, in order to compare the latent scales, a latent space for locating the extracted latent classes is constructed by a technique similar to canonical analysis. Numerical examples are also presented to illustrate the present analysis.

1. Introduction

Scalogram analysis proposed by Guttman (1950) is a method for ordering the subjects measured by using several binary items. In this analysis, the definiteness of responses to items is assumed, and the model for the analysis is as follows. Now, let $X_i$ be a binary item which is an indicator of characteristic $S_i$ ($i = 1, 2, \cdots, n$) to be measured, and suppose that the characteristics $S_i$ are ordered by difficulty with respect to a common trait, i.e., $S_1, S_2, \cdots, S_n$. In this setup, the items are ordered from $X_1$ to $X_n$ in accordance with the orders of the characteristics, and the response patterns to be observed are assumed to be the following $n+1$ patterns:

$$(0, 0, \cdots, 0), (1, 0, \cdots, 0), \cdots, (1, 1 \cdots, 1).$$

This deterministic model is Guttman's perfect scale model. When the above situation holds, Guttman mentined that the trait was "scalable", and the scale was called the Guttman scale or Guttman's perfect scale. Although the model is reasonable, in the practical situation the observed response patterns are obtained beyond the above cases. An improved version of the perfect scale model is the latent distance model which first appeared in Lazarsfeld (1950), and a general discussion of the model is included in Lazarsfeld & Henry (1968). This model is a probabilistic model which admits response errors, and is a restricted case of the latent class model proposed by Lazarsfeld (1950). In this model, the Guttman scale is assumed in latent classes to be extracted. Here, the scale is called the latent Guttman scale. A restricted version of the latent distance model was applied to a probabilistic formulation for Guttman scaling by Proctor (1970), and the method of maximum likelihood (ML) was employed for parameter estimation. In this model, intrusion error rates and omission error rates are the same for all the items to be used. The model may not be realistic in a strict sense, but the study is significant as a new approach to estimation of the parameters. The latent distance model was structured by...

Key Words and Phrases: Guttman's perfect scale model, hierarchical structure, latent class model, latent space, scaling model, scalogram analysis

* Faculty of Engineering, Kyushu Tokai University, Kumamoto, 862 Japan
Eshima & Asano (1988), and they presented an algorithm for ML estimation of the parameters by use of the EM algorithm (Dempster, et al., 1977). By the procedure, the latent distance analysis can now be easily applied to the practical data analysis.

In many scientific fields, the response structures are more complex than the latent distance model, and in such cases the observed subjects cannot be ordered according to one latent scale. For this reason, several scaling models have been proposed, e.g., Dayton & Macready (1976, 1980), Goodman (1975), Eshima & Asano (1988) and Eshima (1990). Eshima (1990) proposed a structured model which is an extension of Eshima-Asano's model (Eshima & Asano, 1988), and presented an ML estimation procedure for the model. In this model, if the algorithm converges, the hypothesized response structure and the set of extracted latent classes can be always identified. In the other models, it is pointed out that the set of estimated latent classes and the hypothesized response structure may not be identified after estimating the parameters, because these models are simple latent class models with equality constraints. In the present paper, the above models are called "scaling models", and the analysis by use of these models "latent scalogram analysis".

The studies with the above models are usually employed in confirmatory analyses. An exploratory case appeared in Price et al. (1980). Their approach was an attempt to select the most suitable hierarchical structure among various structures. This approach is significant as a discovery analysis of hierarchical structures.

The aim of the present paper is to present a general discussion of latent scalogram analysis. As a model Eshima's model (Eshima, 1990) is employed for extracting a hierarchical structure from the data, and some statistical methods for the analysis are considered. Numerical examples are also presented to illustrate the present approach.

2. Scaling models

In this section, scaling models are reviewed briefly, and are compared. Let item $X_i$ be an indicator of acquisition of skill or characteristic $S_i (i = 1, 2, \ldots, n)$. In the present paper, "skill" is used for convenience. In this case, the skills $S_i$ are regarded as latent binary variables. It is assumed that "$S_i = 0$" represents the latent state of non-acquisition of the skill, and that "$S_i = 1$", the state of acquisition of it. Many of the scaling models are in the context of the latent class model proposed by Lazarsfeld (1950), and the latent response probability to each item $X_i$ is assumed to depend on the value of $S_i$. Let $v (s_1, s_2, \ldots, s_n)$ be the proportion of individuals with skill acquisition pattern $(S_1, S_2, \ldots, S_n) = (s_1, s_2, \ldots, s_n)$, and let $Q$ be a set of patterns $(s_1, s_2, \ldots, s_n)$. The patterns are called latent patterns in the present paper. Let $p (x_1, x_2, \ldots, x_n)$ be the manifest probability with which an individual in a population responds with $(X_1, X_2, \ldots, X_n) = (x_1, x_2, \ldots, x_n)$. Then the manifest probability is expressed as

$$p(x_1, x_2, \ldots, x_n) = \sum_{Q} v(s_1, s_2, \ldots, s_n) \prod_{i=1}^{n} P(X_i = 1 | S_i = s_i) P(X_i = 0 | S_i = s_i)^{1-x_i},$$

where $P(\cdot | S_i = s_i)$ represents the conditional probability given $S_i = s_i$, and $\sum_{Q}$ implies the summation over all the latent patterns $(s_1, s_2, \ldots, s_n)$ in $Q$. An improved version of Guttman's perfect scale model is the latent distance model (Lazarsfeld & Henry, 1968), and is given as follows:
\[ P(X_i = 1 \mid S_i = s_i) = \begin{cases} a_i & (s_i = 0) \\ b_i & (s_i = 1) \end{cases} \]

(2.1)

and

\[ 0 \leq a_i \leq b_i \leq 1, \]

(2.2)

in this model, \( a_i \) indicates the intrusion error rate and \( 1 - b_i \) the omission error rate. In this case, a linear hierarchical structure is assumed, and the latent sample space of \((S_1, S_2, \ldots, S_n)\) is as follows:

\[ \mathcal{Q} = \{(0, 0, \ldots, 0), (1, 0, \ldots, 0), \ldots, (1, 1, \ldots, 1)\}. \]

In the present paper, the latent patterns as above are identified to the latent classes. Proctor (1970) applied a restricted model of the above case to explain a linear hierarchical structure. Proctor's model is as follows:

\[ P(X_i = 1 \mid S_i = s_i) = \begin{cases} a & (s_i = 0) \\ 1 - a & (s_i = 1) \end{cases} \]

(2.3)

In this model, the rates of intrusion errors and omission errors are the same for all the items, and the rate is \( a \). The restriction was imposed on the latent distance model because of the difficulty of estimation of the parameters.

In general, restriction (2.2) in the latent distance model was not taken into account in the estimation of the parameters. For this reason improper solutions often occurred. The improper solutions imply not only meaningless estimates of latent probabilities such as \( b_i < 0, a_i > 1 \) and so on, but also the following estimates: \( 0 \leq b_i < a_i \leq 1 \). In these cases the estimated latent classes cannot be identified to the hypothesized structure \( \mathcal{Q} \).

Dayton & Macready (1976) proposed a model for explaining a general hierarchical structure which allows intrusion and omission errors. The model is represented as follows:

\[ P(X_i = 1 \mid S_i = s_i) = \begin{cases} \beta_i & (s_i = 0) \\ 1 - \beta_i & (s_i = 1) \end{cases} \]

(2.4)

Here the model is called Dayton-Macready's model (I). In this model, the parameter \( \alpha \) indicates the omission error rate and \( \beta \) the intrusion error rate. The omission error rate and intrusion error rate are constant through items. Moreover, Dayton & Macready (1980) proposed the following model:

\[ P(X_i = 1 \mid S_i = s_i) = \begin{cases} \beta_i & (s_i = 0) \\ 1 - \beta_i & (s_i = 1) \end{cases} \]

(2.5)

This model is called Dayton-Macready's model (II) in the present paper. In this model, the rate of omission error is the same as that of intrusion error in terms of each item. Concerning the analyses based on these models (2.4) and (2.5), Dayton & Macready (1980) commented as follows:

Models more complex than those models can be set up, but empirical investigation suggested that most are not identified for case of interest. For example, allowing
unrestricted intrusion and omission error rate per item for a Guttman scale yields a model which is not identified.

Eshima & Asano (1988) proposed a structured model for latent distance analysis and a procedure for ML estimation of the parameters. Moreover Eshima (1990) developed the model into a general model for analyzing not only linear hiererchical structures but also branching structures. The model is given as follows:

\[
P(X_i = 1 | S_i = s_i) = \begin{cases} 
\frac{\exp(a_i)}{1 + \exp(a_i)} & (s_i = 0) \\
\frac{\exp(a_i + b_i)}{1 + \exp(a_i + b_i)} & (s_i = 1),
\end{cases}
\]

where

\[b_i = \exp(\beta_i).\]

The parameter \(b_i\) indicates the effect of acquisition of skill \(S_i\) on the response to item \(X_i\). In this model the following inequalities hold true.

\[0 < P(X_i = 1 | S_i = 0) < P(X_i = 1 | S_i = 1) < 1.\]  

Eshima's model is a new model for explaining a general hierarchiical structure. The model allows unrestricted intrusion and omission error rates per items. If the ML estimation algorithm proposed by Eshima (1990) converges, according to (2.7) the estimated model can be always identified to the hypothesized latent structure. The results which are reported in Dayton & Macready (1980) cannot be derived in the analysis using Eshima's model.

3. Model selection procedures

It is important to consider model selection procedures in order to extract a suitable hierarchiical structure. In this section, model selection procedures are mentioned in both exploratory and confirmatory contexts.

3.1. Model selection procedures in exploratory case

Exploratory analysis was discussed in Price et al. (1980). They used Dayton-Macready's model (I), and proposed the backward elimination algorithm (BKELA) and the stepwise forward selection algorithm (SFSA). In these procedures, to select the most suitable model among various models, the likelihood ratio criterion is employed. In this criterion, it is not easy to formally decide the level of significance for performing the likelihood ratio test, so the model to be selected depends on the level to be used. In the present paper, for the above procedures the information criterion AIC (Akaike, 1988) is used in order to select the most suitable model instead of the likelihood ratio test statistic, and as a model Eshima's model is employed. The information criterion AIC is defined as follows:

\[AIC = -2 \times (\text{log maximum likelihood}) + 2 \times (\text{the number of parameters}).\]

In this case, the stopping rule is as follows. Let the values of AIC of the models in Stage \(k\) and Stage \(k+1\) be \(AIC_k\) and \(AIC_{k+1}\), respectively. If \(AIC_k\) is less than \(AIC_{k+1}\), the
algorithm is stopped, and the model in Stage $k$ is selected as the most suitable model for the data set.

3.2. Model selection procedure in confirmatory case

In the confirmatory case, the existence of several latent scales are assumed in the population concerned before analyzing the data. The model selection is performed under this assumption. Suppose that there exist at most the following three latent scales in the population.

\[
\begin{align*}
\text{Scale 1:} & \quad S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_4, \\
\text{Scale 2:} & \quad S_4 \rightarrow S_2 \rightarrow S_3 \rightarrow S_1, \\
\text{Scale 3:} & \quad S_3 \rightarrow S_1 \rightarrow S_2 \rightarrow S_4.
\end{align*}
\]

From these scales the following models are derived.

- Model 1: the above three scales are assumed to be exist,
- Model 2: Scale 1 and Scale 2 are assumed to be exist,
- Model 3: Scale 2 and Scale 3 are assumed to be exist,
- Model 4: Scale 1 and Scale 3 are assumed to be exist,
- Model 5: Scale 1 is assumed to be exist,
- Model 6: Scale 2 is assumed to be exist,
- Model 7: Scale 3 is assumed to be exist.

In this case the model with the minimum value of AIC is selected as the most suitable model among the above models.

In the above discussion, the model selection procedures are formally used for extracting a suitable model. On the other hand, if some information or knowledge about the phenomenon under consideration is available, the model to be used for analyzing the data should be selected by use of that information or knowledge.

4. A dynamic interpretation of latent scales

In this section, we consider a dynamic interpretation of latent scales. A dynamic interpretation of latent scales was given in Eshima (1990) in order to explain a hierarchical learning structure. Here the dynamic interpretation is considered through a mathematical viewpoint. First the case of a linear hierarchical structure is discussed. For simplifying the discussion, it is assumed that there exists the following latent scale:

\[ S_1 \rightarrow S_2 \rightarrow \cdots \rightarrow S_{n-1} \rightarrow S_n. \]  

Now, let us consider the following conditional probability:

\[ P(S_i = s_i | (S_1, S_2, \cdots, S_{i-1}) = (s_1, s_2, \cdots, s_{i-1})). \]

In this setting, if $S_{i-1} = 0$, it means that

\[ (S_1, S_2, \cdots, S_{i-1}) = (1, 1, \cdots, 1) \]

with probability 1, and if $S_{i-1} = 0$, it indicates that

\[ S_i = 0 \]
with probability 1. Thus the conditional probability (4.2) becomes as follows:
\[
P\{S_i=s_i \mid (S_1, S_2, \ldots, S_{i-1})=(s_1, s_2, \ldots, s_{i-1})\} = \begin{cases} 
0 & (S_{i-1}=0 \text{ and } S_i=1) \\
1 & (S_{i-1}=0 \text{ and } S_i=0) 
\end{cases} \quad (4.3)
P(S_i=s_i \mid S_{i-1}=1) \quad (S_{i-1}=1).
\]

From (4.3) we see that the conditional probability depends only on the state of \( S_{i-1} \), so we have the following theorem.

**Theorem 4.1.** *Latent scale (4.1) is a Markov chain.*

This is an interesting result. Let
\[
P(S_i=s_i \mid S_{i-1}=s_{i-1}) = m_{s_i-1,s_i-1}.
\]

Since \( m_{00,i-1}=1 \) and \( m_{01,i-1}=0 \), transition matrix \( M_{i-1} \) from \( S_{i-1} \) to \( S_i \) is given as follows:
\[
M_{i-1} = \begin{pmatrix} 0 & 1 \\
1 & 0 & m_{10,i-1} \\
m_{11,i-1} 
\end{pmatrix} \quad (S_{i-1}).
\]

Although \( S_i \)'s are not sequentially observed at time points, theorem 4.1 induces a dynamic interpretation of latent scale (4.1). The transition probability \( m_{11,i-1} \) may imply the intensity of acquisition of skill \( S_i \) after acquiring skill \( S_{i-1} \). Moreover we get the following theorem.

**Theorem 4.2.** *The transition probabilities \( m_{10,i-1} \) and \( m_{11,i-1} \) can be expressed as follows:
\[
m_{10,i-1} = \frac{v(1, \ldots, 1, 0, \ldots, 0)}{v(1, \ldots, 1, 0, \ldots, 0)+v(1, \ldots, 1, 0, \ldots, 0)+\ldots+v(1,1,1,1)} \quad (4.4)
\]
\[
m_{11,i-1} = \frac{v(1, \ldots, 1, 0, \ldots, 0)+v(1, \ldots, 1, 0, \ldots, 0)+\ldots+v(1,1,1,1)}{v(1, \ldots, 1, 0, \ldots, 0)+v(1, \ldots, 1, 0, \ldots, 0)+\ldots+v(1,1,1,1)} \quad (4.5)
\]

Proof.
\[
m_{10,i-1} = P(S_i=0 \mid S_{i-1}=1) = \frac{P(S_i=0, S_{i-1}=1)}{P(S_{i-1}=1)}. \quad (4.6)
\]

On the other hand, we have
\[ P(S_{i-1}=1) = P\{(S_1, S_2, \ldots, S_{i-1})=(1, 1, \ldots, 1), (S_i, S_{i+1}, \ldots, S_n)=(0, 0, \ldots, 0)\} + P\{(S_1, S_2, \ldots, S_i)=(1, 1, \ldots, 1), (S_{i+1}, S_{i+2}, \ldots, S_n)=(0, 0, \ldots, 0)\} + \cdots + P\{(S_1, S_2, \ldots, S_n)=(1, 1, \ldots, 1)\} \]

\[ = v(1, \ldots, 1, 0, \ldots, 0) + v(1, \ldots, 1, 0, \ldots, 0) + \cdots + v(1, 1, \ldots, 1), \quad (4.7) \]

and

\[ P(S_i=0, S_{i-1}=1) = P\{(S_1, \ldots, S_{i-1})=(1, 1, \ldots, 1), (S_i, \ldots, S_n)=(0, 0, \ldots, 0)\} \]

\[ = v(1, \ldots, 1, 0, \ldots, 0). \quad (4.8) \]

Thus formula (4.4) follows.

Concerning formula (4.5), we have

\[ m_{11,1} = 1 - m_{10,1}. \quad Q.E.D. \]

Second, a general case is discussed. Suppose that there exist the three latent scales given by (3.1). The scales are represented in terms of latent patterns \((s_1, s_2, s_3, s_4)\):

\[
\begin{align*}
    \begin{pmatrix} 1, 0, 0, 0 \\ 0, 0, 0, 0 \\ 0, 0, 1, 0 \end{pmatrix} & \rightarrow \begin{pmatrix} 1, 1, 0, 0 \\ 0, 1, 1, 0 \end{pmatrix} \\
    \begin{pmatrix} 1, 0, 1, 0 \\ 0, 0, 1, 0 \end{pmatrix} & \rightarrow \begin{pmatrix} 1, 1, 1, 0 \end{pmatrix}
\end{align*}
\]

(4.9)

Let \(v(s_1, s_2, s_3, s_4, \text{Scale } j)\) be the proportion of individuals with latent pattern \((s_1, s_2, s_3, s_4)\) in latent scale \(j, j=1, 2, 3\). Then

\[ v(s_1, s_2, s_3, s_4) = \sum_{j=1}^{3} v(s_1, s_2, s_3, s_4, \text{Scale } j). \]

In (4.9), the following equations hold.

\[
\begin{align*}
v(0, 0, 0, 0) &= v(0, 0, 0, 0, \text{Scale 1}) + v(0, 0, 0, \text{Scale 2}) + v(0, 0, 0, \text{Scale 3}) , \\
v(1, 1, 0, 0) &= v(1, 1, 1, 0, \text{Scale 1}) + v(1, 1, 1, \text{Scale 2}) + v(1, 1, 1, \text{Scale 3}) , \\
v(1, 1, 1, 1) &= v(1, 1, 1, 1, \text{Scale 1}) + v(1, 1, 1, 1, \text{Scale 2}) + v(1, 1, 1, 1, \text{Scale 3}) , \quad (4.10) \\
v(0, 0, 1, 0) &= v(0, 0, 1, 0, \text{Scale 2}) + v(0, 0, 1, 0, \text{Scale 3}) , \\
v(1, 0, 0, 0) &= v(1, 0, 0, 0, \text{Scale 1}) , \\
v(1, 1, 0, 0) &= v(1, 1, 0, 0, \text{Scale 1}) , \\
v(1, 0, 1, 0) &= v(1, 0, 1, 0, \text{Scale 2}) , \\
v(1, 0, 1, 1) &= v(1, 0, 1, 1, \text{Scale 3}) .
\end{align*}
\]

In the above setup, each sequence in (3.1) is a Markov chain of its own, but in the above equations (4.10), the parameters \(v(s_1, s_2, s_3, s_4)\) and \(v(s_1, s_2, s_3, s_4, \text{Scale } j)\) are not identified, so we need to impose a restriction on the parameters. In the next section, a method for evaluation of mixed rates of latent scales is presented.
5. Evaluation of mixed rates of latent scales

In the case where there exist several latent scales in the population, the population is divided into several groups according to the scales. It is important to evaluate the proportions of the groups in the population. Suppose that there are three latent scales presented in (3.1). The scales are illustrated in (4.9) by use of the latent patterns \((s_1, s_2, s_3, s_4)\). Let \(w_j\) be the proportion of the group according to Scale \(j\) \((j=1, 2, 3)\). Then we have

\[
w_1 + w_2 + w_3 = 1. \tag{5.1}
\]

In (4.9) we can say that the individuals with latent pattern \((1, 0, 0, 0)\) or \((1, 1, 0, 0)\) depend on Scale 1, and similarly that the latent patterns \((0, 1, 1, 0)\) and \((1, 0, 1, 0)\) come from Scale 2 and Scale 3, respectively. Nevertheless concerning the other latent patterns, we cannot definitely say which of the above three scales they are generated from. Therefore in the present paper the following assumption is imposed on the population. When a latent pattern is derived from several latent scales, it is assumed that the rates of individuals with the pattern which is derived from each scale is in proportion to the mixed rates of the scale in the population. For example, in the above case, the individuals with the latent pattern \((0, 0, 1, 0)\) depend on Scale 2 or Scale 3, so the rates of the individuals with the pattern who depend on Scale 2 and Scale 3 are proportional to \(w_2\) and \(w_3\), respectively, i.e.,

\[
v(0, 0, 1, 0, \text{Scale 2}) = \frac{w_2}{w_2 + w_3} v(0, 0, 1, 0) \]

and

\[
v(0, 0, 1, 0, \text{Scale 3}) = \frac{w_3}{w_2 + w_3} v(0, 0, 1, 0). \]

Under the above assumption, for Scale 1 we have

\[
\begin{align*}
v(0, 0, 0, 0, \text{Scale 1}) &= \frac{w_1}{w_1 + w_2 + w_3} v(0, 0, 0, 0) = w_1 v(0, 0, 0, 0), \\
v(1, 0, 0, 0, \text{Scale 1}) &= v(1, 0, 0, 0), \\
v(1, 1, 0, 0, \text{Scale 1}) &= v(1, 1, 0, 0), \\
v(1, 1, 1, 0, \text{Scale 1}) &= w_1 v(1, 1, 1, 0), \\
v(1, 1, 1, 1, \text{Scale 1}) &= w_1 v(1, 1, 1, 1). \\
\end{align*} \tag{5.2}
\]

Since

\[
w_1 = v(1, 0, 0, 0, \text{Scale 1}) + v(1, 1, 0, 0, \text{Scale 1}) \\
+ v(1, 1, 1, 0, \text{Scale 1}) + v(1, 1, 1, 1, \text{Scale 1}),
\]

we get

\[
w_1 = w_1 \{v(0, 0, 0, 0) + v(1, 1, 1, 0) + v(1, 1, 1, 1)\} \\
+ v(1, 0, 0, 0) + v(1, 1, 0, 0). \tag{5.3}
\]

Similarly, we obtain
\[
W_2 = W_2 \{ v(0, 0, 0, 0) + v(1, 1, 1, 0) + v(1, 1, 1, 1) \} \\
+ \frac{W_2}{W_2 + W_3} v(0, 0, 1, 0) + v(1, 1, 1, 0),
\]
(5.4)

and
\[
W_3 = W_3 \{ v(0, 0, 0, 0) + v(1, 1, 1, 0) + v(1, 1, 1, 1) \} \\
+ \frac{W_3}{W_2 + W_3} v(0, 0, 1, 0) + v(1, 0, 1, 0).
\]
(5.5)

From the above equations we get
\[
\begin{align*}
W_1 &= v(1, 0, 0, 0) + v(1, 1, 0, 0) \frac{1 - \{ v(0, 0, 0, 0) + v(1, 1, 1, 0) + v(1, 1, 1, 1) \}}{v(0, 1, 1, 0)}, \\
W_2 &= (1 - W_1) \times \frac{v(0, 1, 1, 0)}{v(0, 1, 1, 0) + v(1, 0, 1, 0)}, \\
W_3 &= 1 - W_1 - W_2.
\end{align*}
\]
(5.6)

By using the above method, we can evaluate the mixed rates of latent scales.

6. Construction of a latent space

When the number of extracted latent classes is greater than two, it is important to construct a latent space for locating them. The extracted latent classes can then be compared according to their locations. In this section, a latent space for locating the extracted latent classes is constructed according to a technique similar to canonical analysis. Suppose that there are \( A \) latent classes. Let \( v_a \) be the proportion of latent class \( a \) (\( a = 1, 2, \ldots, A \)), and let \( p_{ia} \) and \( p_i \) be the positive response probabilities to item \( X_i \) (\( i = 1, 2, \ldots, n \)) in latent class \( a \) and in the population, respectively. Now, suppose that a new score for the response \( X = (X_1, X_2, \ldots, X_n)' \) is given as
\[
T = \sum_{i=1}^{n} \left( y_{i0} (1 - X_i) + y_{i1} X_i \right),
\]
(6.1)
where \( y_{i0} \) and \( y_{i1} \) are the weights for responses to \( X_i \). The objective in the present discussion is to determine \( y_{ij} \)'s which express the differences among latent classes. Let
\[
Z_i = y_{i0} (1 - X_i) + y_{i1} X_i, \quad i = 1, 2, \ldots, n.
\]
In this setup we have
\[
V(Z_i) = (y_{i0} - y_{i1})^2 p_i (1 - p_i),
\]
(6.2)
and
\[
\text{Cov}(Z_i, Z_j) = (y_{i0} - y_{i1})(y_{j0} - y_{j1}) \sum_{a=1}^{A} v_a (p_{ia} - p_i)(p_{ja} - p_j), \quad i \neq j,
\]
(6.3)
From (6.2) and (6.3), we see that \( y_{i0} \) and \( y_{i1} \) are not identified, so in order to avoid the indeterminacy we put \( y_{i0} = 0 \) and \( y_{i1} = y_i \). Let
\[
\Sigma_B = (\sigma_{ijB}) \quad \text{and} \quad \Sigma_W = (\sigma_{ijW}),
\]
where
\[
\sigma_{ijB} = \sum_{a=1}^{\delta} v_a (p_{ia} - \bar{p}_i)(p_{ja} - \bar{p}_j),
\]
and
\[
\sigma_{ijw} = \begin{cases} 
\sum_{a=1}^{\delta} v_a p_{ia}(1 - p_{ia}) & (i = j) \\
0 & (i \neq j).
\end{cases}
\]

From this setup, we have
\[
V(T) = (y_1, y_2, \ldots, y_n) \Sigma_B (y_1, y_2, \ldots, y_n) + (y_1, y_2, \ldots, y_n) \Sigma_W (y_1, y_2, \ldots, y_n). \tag{6.4}
\]

The first term of the right hand side of equation (6.4) represents the between-class variance of \( T \), and the second term indicates the within-class variance of \( T \). Let
\[
V_B(T) = (y_1, y_2, \ldots, y_n) \Sigma_B (y_1, y_2, \ldots, y_n) \text{'}
\]
and let
\[
V_W(T) = (y_1, y_2, \ldots, y_n) \Sigma_W (y_1, y_2, \ldots, y_n) \text{'}.
\]

In order to determine the first weight vector \( y = (y_1, y_2, \ldots, y_n) \text{'} \), which expresses the differences of latent classes, the following criterion is employed:
\[
\frac{V_B(T)}{V_W(T)} \rightarrow \text{Max}. \tag{6.5}
\]

In this criterion, to avoid the indeterminacy with respect to the scales of \( y_i \)'s, we impose the following constraint on \( y_i \)'s.
\[
V_W(T) = (y_1, y_2, \ldots, y_n) \Sigma_W (y_1, y_2, \ldots, y_n) = 1. \tag{6.6}
\]

Under the assumption of (6.6), (6.5) is equivalent to
\[
V_B(T) \rightarrow \text{Max}. \tag{6.7}
\]

In order to obtain \( y \) we define the following function:
\[
f(y) = y \Sigma_B y' + \lambda (1 - y \Sigma_W y'). \tag{6.8}
\]

Differentiating the above function with respect to \( y \) and setting it to zero, we get
\[
\frac{\partial f}{\partial y} = 2 \Sigma_B y - 2 \lambda \Sigma_W y = 0. \tag{6.9}
\]

If \( \Sigma_W \) is not singular, from (6.9) we have
\[
(\Sigma^{-1/2} \Sigma_B \Sigma^{-1/2} - \lambda I) \Sigma_{1/2} y = 0, \tag{6.10}
\]
where \( I \) is the identity matrix of order \( n \). As \( y \neq 0 \), we see that \( \lambda \) is an eigenvalue of \( \Sigma^{-1/2} \Sigma_B \Sigma^{-1/2} \). Suppose that the rank of the matrix \( \Sigma_B^{2/2} \Sigma_B \Sigma_B^{2/2} \) is \( k \) and that the matrix has \( k \) different eigenvalues. Let \( \lambda_i \) be the \( i \)-th largest eigenvalue and let \( \xi_i \) be the eigenvector related to \( \lambda_i \) (\( i = 1, 2, \ldots, k \)). When we put \( y_i = \Sigma_W^{-1/2} \xi_i \), we have
\[
V_B(T) = y_i \Sigma_B y_i = \xi_i \Sigma_B^{2/2} \Sigma_B \Sigma_B^{2/2} \xi_i = \lambda_i. \tag{6.11}
\]

From (6.11) we see that \( y_i \) is the solution to (6.7). Thus \( T_1 = y'X \) represents the first axis.
which explains the maximum variation of latent classes in the sense of criterion (6.5). From a similar discussion of canonical analysis, the \( l \)-th weight vector, which expresses the differences of latent classes as large as possible with respect to variance of \( T_i = y_i X \) and which is orthogonal to \( y_m (m = 1, 2, \ldots, l - 1) \) with respect to \( \Sigma_w \), i.e., \( y' \Sigma_w y_m = 0 \), can be obtained as

\[
y_i = \Sigma_w^{-1/2} \xi_i.
\]

Thus by induction, we can derive \( k \) axes: \( T_1, T_2, \ldots, T_k \). Concerning the contributions of derived axes to express the differences of extracted latent classes, the orders are those of the magnitudes of \( \lambda_i \)'s. Hence a suitable number of axes can be determined according to the contributions. In the present case, it may be suitable to select two or three axes. The above discussion derives the same technique used in Ikuzawa (1968).

Let the number of the selected axes be \( r \). The extracted latent classes can be located in the \( r \)-dimensional Euclidian space constructed by the above discussion. Let

\[
p_a = (p_{1a}, p_{2a}, \ldots, p_{na})'.
\]

Then the location of latent class \( a \) is defined as \((t_{1a}, t_{2a}, \ldots, t_{ra})\), where

\[
t_{ia} = E(T_i | a) = y_i' p_a.
\]

By use of the above latent space, the latent classes and latent scales can be compared.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Stouffer-Toby's data set, McHugh's data set and Lazarsfeld-Stouffer's data set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set 1</td>
<td>Data Set 2</td>
</tr>
<tr>
<td>( X_1, X_2, X_3, X_4 )</td>
<td>Freq.</td>
</tr>
<tr>
<td>0 0 0 0</td>
<td>42</td>
</tr>
<tr>
<td>1 0 0 0</td>
<td>1</td>
</tr>
<tr>
<td>0 1 0 0</td>
<td>6</td>
</tr>
<tr>
<td>1 1 0 0</td>
<td>2</td>
</tr>
<tr>
<td>0 0 1 0</td>
<td>6</td>
</tr>
<tr>
<td>1 0 1 0</td>
<td>1</td>
</tr>
<tr>
<td>0 1 1 0</td>
<td>7</td>
</tr>
<tr>
<td>1 1 1 0</td>
<td>2</td>
</tr>
<tr>
<td>0 0 0 1</td>
<td>23</td>
</tr>
<tr>
<td>1 0 0 1</td>
<td>4</td>
</tr>
<tr>
<td>0 1 0 1</td>
<td>24</td>
</tr>
<tr>
<td>1 1 0 1</td>
<td>9</td>
</tr>
<tr>
<td>0 0 1 1</td>
<td>25</td>
</tr>
<tr>
<td>1 0 1 1</td>
<td>6</td>
</tr>
<tr>
<td>0 1 1 1</td>
<td>38</td>
</tr>
<tr>
<td>1 1 1 1</td>
<td>20</td>
</tr>
</tbody>
</table>
7. Numerical examples

In this section, the present methods are demonstrated by use of 4 data sets.

7.1. Exploratory case

Three data sets are shown in Table 1, and these data sets were used for demonstrating BKELA and SFSA by Price, et al. (1980). In Price, et al. (1980) the details of these data sets are mentioned as follows:

Data set 1 is Stouffer-Toby's data from respondents to questionnaire items on role conflict. This data represents a cross-classification of respondents with respect to whether they tend toward "universalistic" values (1) or "particularistic" values (0) when confronted by each of four different situations of role conflict.

Data set 2 is McHugh's test data on creative ability in machine design. This data represents a cross-classification of engineers with respect to their dichotomized scores, i.e., above the subtest mean (1) or below (0), obtained on each of four different subtests that measured creative ability in machine design.

Data set 3 is Lazarsfeld-Stouffer's data from noncommissioned officers responding to items on attitude toward the Army. This data represents a cross-classification of a sample of respondents with respect to their dichotomized responses, i.e., favorable (1) or unfavorable (0), obtained on each of four different items on general attitude toward the Army.

In the present demonstration, Eshima's model was employed in BKELA and SFSA, and the information criterion AIC was used for comparing the goodness-of-fits of the models. Because of the necessary condition for model identification, the number of latent classes to be extracted is less than or equal to 8 for the present data sets. In general, the condition is as follows:

the number of latent classes \( \leq 2^n - 2n \),

where \( n \) is the number of items \( X_i \). Thus in Stage 1 of BKELA the following latent patterns \( (s_1, s_2, s_3, s_4) \) were hypothesized for each data set by considering the frequencies of response patterns in the data set.

Stouffer-Toby's data set: \((0, 0, 0, 0), (0, 1, 1, 0), (0, 0, 1, 0), (0, 1, 0, 1), (1, 1, 0, 1), (0, 0, 1, 1), (0, 1, 1, 1), (1, 1, 1, 1)\).

McHugh's data set: \((0, 0, 0, 0), (1, 0, 0, 0), (0, 1, 0, 0), (1, 1, 0, 0), (1, 1, 1, 0), (0, 0, 1, 1), (0, 1, 1, 1), (1, 1, 1, 1)\).

Lazarsfeld-Stouffer's data set: \((0, 0, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1), (0, 1, 0, 1), (0, 0, 1, 1), (1, 0, 1, 1), (0, 1, 1, 1), (1, 1, 1, 1)\).

If we need to increase the number of latent classes to be extracted, Eshima's model must be restricted. For example, \( a_i = a \) (for all \( i, i=1, 2, \ldots, n \)) or \( b_i = b \) (for all \( i, i=1, 2, \ldots, n \)). In SFSA, the latent patterns in the first stage are \((0, 0, 0, 0)\) and \((1, 1, 1, 1)\). In both algorithms the results were the same for each data set. The results of the analyses of the above data sets are shown in Table 2. On the other hand, for data set 1 and 3 the results in Price et al. (1980) depended on the given levels of significance in the likelihood.
Table 2
The Results of the Analyses of the Three Data Sets in Table 1
by use of BKELA and SFSA

Stouffer-Toby's Data Set

<table>
<thead>
<tr>
<th>Latent Patterns</th>
<th>Class Proportions</th>
<th>Item Positive Response Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 0, 0, 0)</td>
<td>0.280</td>
<td>0.007 0.061 0.074 0.232</td>
</tr>
<tr>
<td>(1, 1, 1, 1)</td>
<td>0.720</td>
<td>0.287 0.671 0.646 0.868</td>
</tr>
</tbody>
</table>

$\chi^2 = 2.720$, d.f. = 6, $P$-val. = 0.843, AIC = 75.801.

McHugh's Data Set

<table>
<thead>
<tr>
<th>Latent Patterns</th>
<th>Class Proportions</th>
<th>Item Positive Response Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 0, 0, 0)</td>
<td>0.396</td>
<td>0.289 0.244 0.112 0.204</td>
</tr>
<tr>
<td>(1, 1, 0, 0)</td>
<td>0.077</td>
<td>0.894 0.996 0.112 0.204</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>0.200</td>
<td>0.239 0.244 0.979 0.827</td>
</tr>
<tr>
<td>(1, 1, 1, 1)</td>
<td>0.327</td>
<td>0.894 0.996 0.979 0.827</td>
</tr>
</tbody>
</table>

$\chi^2 = 1.100$, d.f. = 4, $P$-val. = 0.894, AIC = 75.134.

Lazarsfeld-Stouffer's Data Set

<table>
<thead>
<tr>
<th>Latent Patterns</th>
<th>Class Proportions</th>
<th>Item Positive Response Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 0, 0, 0)</td>
<td>0.280</td>
<td>0.007 0.061 0.074 0.232</td>
</tr>
<tr>
<td>(1, 1, 1, 1)</td>
<td>0.720</td>
<td>0.287 0.671 0.646 0.868</td>
</tr>
</tbody>
</table>

$\chi^2 = 8.525$, d.f. = 6, $P$-val. = 0.202, AIC = 105.009.

ratio test and the methods, BKELA and SFSA. For data set 2, the obtained latent patterns in Table 2 are the same as those in Price et al. (1980). The goodness-of-fit of the selected model for each data set is better than that of Price et al. (1980). For Stuffer-Toby's data set and Lazarsfeld-Stouffer's data set, the extracted latent classes are those with latent patterns (0, 0, 0, 0) and (1, 1, 1, 1). According to the results, we may say that concerning Stuffer-Toby's data set, there are latent "universalistic" and "particularistic" states, and that in Lazarsfeld-Stouffer's data there are two latent states, "favorable" and "unfavorable" states toward the Army. Thus in both cases the latent scale is

(0, 0, 0, 0) → (1, 1, 1, 1).

In the above two cases, it means that $S_1 = S_2 = S_3 = S_4$ and it is not necessary to construct the latent space proposed here.

For McHugh's data set, the following latent scales were extracted.

(0, 0, 0) → (1, 1, 0, 0) → (1, 1, 1, 1).
Fig. 1 Locations of Latent Classes (McHugh's Data Set)

Table 3
Axes of the Latent Space and Locations of the Extracted Latent Classes (McHugh's Data Set)

<table>
<thead>
<tr>
<th>Eigenvalues</th>
<th>Scores</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>$y_1$</td>
<td>$y_2$</td>
<td>$y_3$</td>
<td>$y_4$</td>
<td></td>
</tr>
<tr>
<td>$T_1$</td>
<td>4.2905</td>
<td>0.7277</td>
<td>1.0986</td>
<td>3.2677</td>
<td>0.9497</td>
</tr>
<tr>
<td>$T_2$</td>
<td>1.2873</td>
<td>1.4017</td>
<td>2.1161</td>
<td>-1.7004</td>
<td>-0.4877</td>
</tr>
<tr>
<td>$T_3$</td>
<td>-0.0002</td>
<td>0.0005</td>
<td>-0.0081</td>
<td>-1.6877</td>
<td>2.3303</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Latent Patterns</th>
<th>Locations of Latent Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T_1$</td>
</tr>
<tr>
<td>(0, 0, 0, 0)</td>
<td>1.034</td>
</tr>
<tr>
<td>(1, 1, 0, 0)</td>
<td>2.304</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>4.426</td>
</tr>
<tr>
<td>(1, 1, 1, 1)</td>
<td>5.729</td>
</tr>
</tbody>
</table>
In this case, it means that
\[ S_1 = S_2 \text{ and } S_3 = S_4, \]
or that \( (S_1, S_2) \) and \( (S_3, S_4) \) are acquired at the same time, respectively. Let
\begin{align*}
\text{Scale 1: } & (0, 0, 0, 0) \rightarrow (1, 1, 0, 0) \rightarrow (1, 1, 1, 1), \\
\text{Scale 2: } & (0, 0, 0, 0) \rightarrow (0, 0, 1, 1) \rightarrow (1, 1, 1, 1).
\end{align*}

Eshima (1990) called the scale learning processes, i.e., he derived a dynamic interpretation from the latent scales. Let \( w_i \) be the proportion of individuals according to Scale \( i \), \( i = 1, 2 \). By the discussion in section 4, we have
\[
\begin{align*}
w_1 + w_2 &= 1, \\
w_1 &= w_1 \{ v(0, 0, 0, 0) + v(1, 1, 1, 1) \} + v(1, 1, 0, 0), \\
w_2 &= w_2 \{ v(0, 0, 0, 0) + v(1, 1, 1, 1) \} + v(0, 0, 1, 1).
\end{align*}
\]

From these equations, we get
\[
\begin{align*}
w_1 &= \frac{v(1, 1, 0, 0)}{v(1, 1, 0, 0) + v(0, 0, 1, 1)}, \\
w_2 &= \frac{v(0, 0, 1, 1)}{v(1, 1, 0, 0) + v(0, 0, 1, 1)}.
\end{align*}
\]

Hence the estimates of \( w_1 \) and \( w_2 \) are given as follows:
\[
\hat{w}_1 = 0.278 \text{ and } \hat{w}_2 = 0.722.
\]

The solution is the same as that of Eshima (1990). The latent space and the location of extracted latent classes are illustrated in Figure 1, and the details for constructing the latent space are presented in Table 3. The first axis implies the general creative ability.
and the second axis expresses the difference between abilities for acquiring skills \((S_1, S_2)\) and \((S_3, S_4)\). By considering the latent space, we can get an overview of the differences among the extracted latent classes, and compare the above two latent scales. For scale 1 the estimated transition matrix is given as follows:

\[
\begin{bmatrix}
0 & 1 \\
0.459 & 0.541
\end{bmatrix}
\]

\(S_3 = S_4\)

\(M_1(\text{scale } 1) =
\begin{bmatrix}
1 & 0 \\
0.459 & 0.541
\end{bmatrix}
\)

\(S_1 = S_2\)

and for scale 2 we get

Table 5  
Latent Structures and Latent Patterns

<table>
<thead>
<tr>
<th>Latent Structure</th>
<th>Latent Patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale (1)</td>
<td>(0, 0, 0, 0)</td>
</tr>
<tr>
<td></td>
<td>(0, 1, 1, 1)</td>
</tr>
<tr>
<td>Scale (2)</td>
<td>(0, 0, 0, 0)</td>
</tr>
<tr>
<td></td>
<td>(0, 1, 1, 1)</td>
</tr>
<tr>
<td>Scale (3)</td>
<td>(0, 0, 0, 0)</td>
</tr>
<tr>
<td></td>
<td>(0, 1, 1, 1)</td>
</tr>
<tr>
<td>Scale (1 • 2)</td>
<td>(0, 0, 0, 0)</td>
</tr>
<tr>
<td></td>
<td>(0, 1, 1, 1)</td>
</tr>
<tr>
<td>Scale (2 • 3)</td>
<td>(0, 0, 0, 0)</td>
</tr>
<tr>
<td></td>
<td>(0, 1, 1, 1)</td>
</tr>
<tr>
<td>Scale (1 • 3)</td>
<td>(0, 0, 0, 0)</td>
</tr>
<tr>
<td></td>
<td>(0, 1, 1, 1)</td>
</tr>
<tr>
<td>Scale (1 • 2 • 3)</td>
<td>(0, 0, 0, 0)</td>
</tr>
<tr>
<td></td>
<td>(0, 1, 1, 1)</td>
</tr>
</tbody>
</table>

Table 6  
Results of the Analyses of Proctor's Data Set

<table>
<thead>
<tr>
<th>Latent Structure</th>
<th>(x^2)-val.</th>
<th>d.f.</th>
<th>(P)-val.</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale (1)</td>
<td>18.353</td>
<td>17</td>
<td>0.367</td>
<td>137.866</td>
</tr>
<tr>
<td>Scale (2)</td>
<td>18.353</td>
<td>17</td>
<td>0.367</td>
<td>137.866</td>
</tr>
<tr>
<td>Scale (3)</td>
<td>18.673</td>
<td>17</td>
<td>0.348</td>
<td>138.186</td>
</tr>
<tr>
<td>Scale (1 • 2)</td>
<td>18.354</td>
<td>16</td>
<td>0.304</td>
<td>139.867</td>
</tr>
<tr>
<td>Scale (2 • 3)</td>
<td>16.500</td>
<td>16</td>
<td>4.419</td>
<td>138.013</td>
</tr>
<tr>
<td>Scale (1 • 3)</td>
<td>16.499</td>
<td>15</td>
<td>0.350</td>
<td>140.012</td>
</tr>
<tr>
<td>Scale (1 • 2 • 3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The obtained matrices are the same.

7.2. Confirmatory case

Table 4 shows the first data set in Proctor (1970). In this data analysis, let us assume that there exist at most the following three scales.
In this case, by combination of the above scales, seven hierarchical structures are derived. For convenience the structures are coded as follows. For example, the structure with Scale 1 and Scale 2 is expressed as Scale (1\cdot2), and the structure with Scale 3 as Scale (3), and so on. The relationships among the structures and the latent patterns to be extracted are shown in Table 5. In Table 5, we see that the sets of latent patterns and hypothesized structures are not completely identified, i.e., Scale (1\cdot2\cdot3) and Scale (1\cdot3) induce the same set of latent patterns. The results of the analysis are included in Table 6. From this table, we see that Scale (1) and Scale (2) show the best fit to the data set according to the AIC, and the estimated latent classes and latent probabilities are shown in Table 7. The results of the test for goodness-of-fit is as follows: \( x^2 = 18.353, \) \( d.f. = 17, \) \( P\text{-val.} = 0.367 \) and \( \text{AIC}=137.866. \) The fit to the data set is good. Which of the two structures is used for this analysis is determined by information or knowledge about the data set. Here
Table 10
Axes of the Latent Space and Locations of the Extracted Latent Classes

<table>
<thead>
<tr>
<th>Eigenvalues</th>
<th>$\lambda_i$</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
<th>$y_4$</th>
<th>$y_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_1$</td>
<td>4.6084</td>
<td>2.1236</td>
<td>0.5310</td>
<td>1.0471</td>
<td>0.4875</td>
<td>2.8840</td>
</tr>
<tr>
<td>$T_2$</td>
<td>0.9142</td>
<td>-2.6073</td>
<td>-0.1931</td>
<td>-0.0426</td>
<td>0.3824</td>
<td>2.9175</td>
</tr>
<tr>
<td>$T_3$</td>
<td>0.3567</td>
<td>-0.9706</td>
<td>0.2612</td>
<td>2.3777</td>
<td>0.8760</td>
<td>-1.5332</td>
</tr>
<tr>
<td>$T_4$</td>
<td>0.1227</td>
<td>0.4603</td>
<td>-0.6176</td>
<td>-0.9696</td>
<td>2.2676</td>
<td>-0.4411</td>
</tr>
<tr>
<td>$T_5$</td>
<td>0.0441</td>
<td>-0.5068</td>
<td>2.1574</td>
<td>-0.8269</td>
<td>0.4574</td>
<td>-0.3894</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Latent Patterns</th>
<th>Locations of Latent Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T_1$</td>
</tr>
<tr>
<td>(0, 0, 0, 0, 0)</td>
<td>1.014</td>
</tr>
<tr>
<td>(0, 0, 0, 1, 0)</td>
<td>1.314</td>
</tr>
<tr>
<td>(0, 0, 1, 1, 0)</td>
<td>2.066</td>
</tr>
<tr>
<td>(0, 0, 0, 0, 1)</td>
<td>3.405</td>
</tr>
<tr>
<td>(0, 0, 0, 1, 1)</td>
<td>3.704</td>
</tr>
<tr>
<td>(0, 0, 1, 1, 1)</td>
<td>4.457</td>
</tr>
<tr>
<td>(0, 1, 1, 1, 1)</td>
<td>4.716</td>
</tr>
<tr>
<td>(1, 1, 1, 1, 1)</td>
<td>6.470</td>
</tr>
</tbody>
</table>

![Fig. 3 Locations of Latent Classes in Table 9 (Proctor's Data Set)](image)

Scale (1) is employed and is discussed. The constructed latent space and the details for constructing the space are shown in Figure 2 and Table 8, respectively. In this case two axes are used for constructing the space. From Figure 2 we see that the latent patterns are divided into the following three groups:

$$G_1 = \{C_1\}, \quad G_2 = \{C_2, C_3, C_4, C_5\}, \quad G_3 = \{C_6\}.$$  

The estimated transition matrices are as follows:
From the above matrices, we get
\[
M_{11,1} = 0.953, \quad M_{11,2} = 0.915, \quad M_{11,3} = 0.948.
\]
These transition probabilities are very high, and hence this may imply that the latent classes, \(C_2, C_3, C_4\) and \(C_5\), consist of a group as shown above.

It is significant to demonstrate an analysis of more complex hierarchical structure than Scale (1). Let us discuss Scale (1 \(\times 2 \times 3\)) below. The estimated latent probabilities are included in Table 9. Table 10 shows the details for constructing the latent space, and the space is illustrated in Figure 3. According to Figure 3, Scale 1 and Scale 2 are very close to each other, but Scale 3 shows a different flow from Scale 1 and 2. This figure may show a reason why Scale (1) and Scale (2) have the same results in Table 6. We see that the extracted latent classes are divided into three groups. The first group consists of \(C_1, C_2\) and \(C_3\), and the individuals in the group do not have Skill 5. The second group is the set of \(C_4, C_5, C_6\) and \(C_7\). The group has Skill 5, but does not have Skill 1. The third group includes only \(C_8\). From this grouping, we may say that Skill 1 and Skill 5 play an important role in this scaling. A more detailed interpretation of the results than the above might be achieved with more practical information or knowledge. As shown above, the latent space considered in this paper is useful for interpreting the latent scales.

Concerning the evaluation of the proportions of latent scales, we obtain
\[
\begin{align*}
\hat{w}_1 &= w_1 \left\{ v(0, 0, 0, 0, 0) + v(0, 0, 1, 1, 1) + v(0, 1, 1, 1, 1) + v(1, 1, 1, 1, 1) \right\} \\
&+ \frac{w_1}{w_1 + w_2} v(0, 0, 0, 1, 1) + v(0, 0, 0, 0, 1), \\
\hat{w}_2 &= w_2 \left\{ v(0, 0, 0, 0, 0) + v(0, 0, 1, 1, 1) + v(0, 1, 1, 1, 1) + v(1, 1, 1, 1, 1) \right\} \\
&+ \frac{w_2}{w_1 + w_2} v(0, 0, 0, 1, 0) + v(0, 0, 1, 1, 0), \\
\hat{w}_3 &= w_3 \left\{ v(0, 0, 0, 0, 0) + v(0, 0, 1, 1, 1) + v(0, 1, 1, 1, 1) + v(1, 1, 1, 1, 1) \right\} \\
&+ \frac{w_3}{w_1 + w_2} v(0, 0, 0, 1, 0) + v(0, 0, 1, 1, 0).
\end{align*}
\]

From these equations we get
\[
\begin{align*}
\hat{w}_1 &= \frac{v(0, 0, 0, 0, 1)}{v(0, 0, 0, 1, 0) + v(0, 0, 0, 0, 1)}, \\
\hat{w}_2 &= 1 - \hat{w}_1 - \hat{w}_3, \\
\hat{w}_3 &= \frac{v(0, 0, 1, 1, 0)}{v(0, 0, 0, 1, 1) + v(0, 0, 1, 1, 0)}.
\end{align*}
\]
Hence the estimates of \(\hat{w}_i\)'s are given as follows:
\[
\hat{w}_1 = 0.246, \quad \hat{w}_2 = 0.368, \quad \hat{w}_3 = 0.386.
\]
8. Conclusion

In the present paper, a general discussion for latent scalogram analysis was presented, and some statistical methods for the analysis were proposed. By use of numerical examples the proposed methods were demonstrated, and the demonstrations showed efficiency of the present approach. In categorical data analysis, it is important to derive a simple structure from the observed data set. The present discussion offers an approach to deriving a simple latent structure based on binary data and to interpreting the extracted structure. In this sense, the present approach is significant.

Acknowledgment

The author wishes to thank the referees and the editor for their useful comments and advice.

References


(Received October, 1991. Revised November, 1991)