Abstract

Although Aitchison’s (Aitchison, J., 1986. The Statistical Analysis of Compositional Data, Chapman and Hall, London) method of logratio transformation of compositional data is widely used in various domains, it is limited by the assumption of a strict non-negativity of the components and the requirement of special treatments in practice of the zero components. We propose a dimension-reduction approach through a hyperspherical transformation that is capable of resolving the difficulty in maintaining non-negativity and unit-sum in forecasting compositional data over time. Applying the proposed model to a numerical simulation with a 4D compositional data embedded with zero components and forecasting the three production sectors in the Chinese economy both demonstrate the usefulness and validity of the new approach.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Forecasting; Data Analysis; Compositional Data Analysis

1. Introduction

We use proportions in the scientific and economic disciplines to measure the shares of the different components of an activity (e.g., Billheimer et al., 2001; Katz and King, 1999). Proportions are useful in the analysis of categorical data (Agresti and Fin-
an economy’s production into primary, secondary, and tertiary sector. Specifically, a compositional data consist of a vector of the $p$ parts of non-negative components $X = (x_1, x_2, \ldots, x_p)$ that represent proportions of some whole subjected to both non-negativity and unit-sum constraints, and is expressed as follows:

$$\sum_{j=1}^{p} x_j = 1, \quad x_j \geq 0. \tag{1}$$

A popular way to represent the data is through a $p$-dimension ternary diagram (although there is no satisfactory visual representation when $p > 4$) with each component $x_i$ corresponding to the perpendicular point from the representative point to the triangular face opposite the vertex $j$.

We investigate the correlation structure or clustering pattern of the components, and hypothesize a systematic trend of the compositional data over time. For example, we identify how much employment will be absorbed by an increasing trend in the tertiary production sector of the economy. However, in the distant past, Pearson (1897) pointed out that a relational analysis on the components of compositional data was prone to obtaining spurious correlations if we ignore or misspecify the unit-sum constraint. Quenouille (1953, 1959) provided the first formal statistical modeling of compositional data. Based on the additive logistic normal distribution, Aitchison (1986) proposed a logratio transformation to the statistical analysis of compositional data. The logratio is defined as $\log(r_i) = \log \left( \frac{x_i}{x_0} \right)$, where $i = 1, \ldots, d$, $x_i = r_i/(r_1 + \cdots + r_d + 1)$, and $x_0 = 1/(r_1 + \cdots + r_d + 1)$ (eqs. (2.6) (2.7) in Aitchison, 1986, p. 26). Fry et al. (2000) extended Aitchison’s (1986) logratio transformation to take into account the zero proportions in micro data. Although Aitchison’s (1986) logratio transformation is widely used in various disciplines, the logarithm of zero is undefined, (e.g., $\log \left( \frac{x_i}{x_0} \right)$ is undefined when $x_i = 0$). He devised special treatments to handle the zero components of compositional data in practice, and lamented that “the problem of zeros is unlikely ever to be satisfactorily and generally resolved (p. 266).”

Multinomial distribution (e.g., Keane, 1992; Geweke et al., 1994; Capon et al., 1990) was developed as a generalization of the binomial distribution by taking into account more than two classes into which an event falls. $\sum_{i} p_i = 1$, where $p_i$ refers to the probability of an individual event falling in the $i$th class, and multinomial distribution is capable of accommodating a zero probability of occurrence. Furthermore, others developed the literature on forecasting (e.g., Klein and Park, 1993) and multivariate time series analysis (e.g., Quintana and West, 1988) to predict future compositional data. However, the proportions are usually restricted to just the positive values to circumvent the violation of the unit-sum constraint in the predicted periods. Maintaining the unit-sum assumption over time in the construction of a forecasting model for compositional data is a difficult problem to resolve.

In this paper, we propose to forecast compositional data through a hyperspherical transformation that resolves the difficulty of zero components and maintains unit-sum in the predicted periods. We derive this dimension-reduction approach from the degrees of freedom of a $p$-dimensional vector as $p - 1$. By applying the proposed model to a numerical simulation with 4-dimensional data embedded with zero components and to forecasting the primary, secondary, and tertiary production compositions in China, we attest to the validity of the new approach for each application. We maintain both the non-negativity and unit-sum constraints in the estimation and prediction periods.

The remainder of our paper is organized as follows. Section 2 presents the dimension-reduction approach through a hyperspherical transformation (DRHT) and develops the algorithm to analyze compositional data. Section 3 reports the simulation that allows for zero components and presents the forecast for the share of primary, secondary, and tertiary productions in China. Section 4 presents the conclusion.

2. The DRHT forecasting model for compositional data

2.1. The issue

Aitchison’s (1986) logratio transformation usually stipulates the strict positive value of every component. Without losing generality, we slightly modify the non-negativity constraint as presented in Eq. (1):

$$\sum_{j=1}^{p} x_j = 1, \quad 0 \leq x_j < 1. \tag{2}$$
Consider a set of compositional data, \( X^t \), that is collected according to the time sequence \( t = 1, 2, \ldots, T \):

\[
X^t = \left\{ (x_1^t, \ldots, x_p^t) \in \mathbb{R}^p \mid \sum_{j=1}^p x_j^t = 1, \ 0 \leq x_j^t < 1 \right\},
\]

\( t = 1, 2, \ldots, T \).

Our objective is to build a model for the given data set \( X^t \) so as to forecast \( X^{T+l} \) at time \( T + l \),

\[
X^{T+l} = \left\{ (x_1^{T+l}, \ldots, x_p^{T+l}) \in \mathbb{R}^p \mid \sum_{j=1}^p x_j^{T+l} = 1, \ 0 \leq x_j^{T+l} < 1 \right\}.
\]

To solve the problem stated above, analysts build models for each element \( x_j^t \) of the vector \( X^t \) by time-series analysis or least-squares regression (e.g., Greene, 1993). That is, for a given \( j = 1, 2, \ldots, p \), the model is built according to the data record \( \{x_j^t, t = 1, 2, \ldots, T\} \), and we subsequently use the model for the forecasting of \( x_j^{T+l} \). However, this kind of modeling often suffers from failing the unit-sum constraint in the predicted periods, i.e., \( \sum_{j=1}^p x_j^{T+l} \neq 1 \).

The reason for the failure of the unit-sum constraint in the predicted periods is simply because the degree of freedom of a p-dimensional vector of the compositional data is only \( p - 1 \). When one builds \( p \) models for each element of the p-dimensional vector, there is a redundant degree of freedom, so that one violates the constraint of the unit-sum.

### 2.2. The DRHT forecasting model

Given the fact that the degree of freedom of a p-dimensional vector is \( p - 1 \), we propose a dimension-reduction approach through a hyperspherical transformation (DRHT) of compositional data. As shown earlier, for a compositional data set, \( X^t \), in \( p \)-dimension at the given time \( t = 1, 2, \ldots, T \),

\[
X^t = \left\{ (x_1^t, \ldots, x_p^t) \in \mathbb{R}^p \mid \sum_{j=1}^p x_j^t = 1, \ 0 \leq x_j^t < 1 \right\},
\]

\( t = 1, 2, \ldots, T \).

DRHT involves a non-linear transformation by mapping a compositional data vector onto a hypersphere in order to reduce its dimension from \( p \) to \( p - 1 \). We present below the modeling procedure and the algorithm of the DRHT, and its illustration on a 3D compositional data is shown in Appendix A.

First, we perform a square root transformation on the elements of the original compositional data.

\[
y_j^t = \sqrt{x_j^t}, \ j = 1, \ldots, p; t = 1, 2, \ldots, T. \tag{4}
\]

Let us denote \( y^t = (y_1^t, \ldots, y_p^t) \) where \( t = 1, 2, \ldots, T \). It is obvious that

\[
||y^t||^2 = \sum_{j=1}^p (y_j^t)^2 = 1. \tag{5}
\]

The vector \( y^t = (y_1^t, \ldots, y_p^t) \in \mathbb{R}^p(t = 1, \ldots, T) \) is on the surface of a p-dimensional hypersphere with radius 1 at any time \( t \).

Second, we map the vector \( y^t = (y_1^t, \ldots, y_p^t) \in \mathbb{R}^p(t = 1, 2, \ldots, T) \) from the Cartesian coordinate system onto a hyperspherical coordinate system \( (r, \theta_2, \ldots, \theta_p)^t \in \Theta^p \). Note that \( (r^t)^2 = ||y^t||^2 = 1 \), and the mapping equations are as follows:

\[
\begin{align*}
    y_1^t & = \sin \theta_1^t \\
    y_2^t & = \cos \theta_1^t \sin \theta_3^t \\
    y_3^t & = \cos \theta_1^t \cos \theta_3^t \\
    & \vdots \\
    y_{p-2}^t & = \cos \theta_{p-2} \sin \theta_{p-1}^t \\
    y_{p-1}^t & = \cos \theta_{p-1} \sin \theta_p^t \\
    y_p^t & = \cos \theta_p^t,
\end{align*}
\]

where \( 0 < \theta_j^t \leq \pi/2, j = 2, 3, \ldots, p \).

In the hyperspherical transformation from Eqs. (4)–(6), the dimension of the vector of the compositional data is reduced from \( p \) to \( p - 1 \). That is, the \( p \) linearly correlated variables have changed into \( p - 1 \) independent angles \( \theta_j^t, j = 2, 3, \ldots, p \). In Eq. (6), a recursive algorithm is applied to compute \( \theta_j^t, j = 2, 3, \ldots, p \).

\[
\begin{align*}
    \theta_1^t & = \arccos y_1^t \\
    \theta_{p-1}^t & = \arccos \frac{y_{p-1}^t}{\sin \theta_p^t} \\
    \theta_{p-2}^t & = \arccos \frac{y_{p-2}^t}{\sin \theta_p^t \sin \theta_{p-1}^t} \\
    & \vdots \\
    \theta_2^t & = \arccos \frac{y_2^t}{\sin \theta_1^t \sin \theta_{p-1}^t \sin \theta_p^t},
\end{align*}
\]

(7)
From the computed angles \( \{\theta_j, t = 1, 2, 3, \ldots, T\} \), \( j = 2, 3, \ldots, p \) in Eq. (7), we build \( p - 1 \) models for each angle data series, respectively. An example of which by a regressive model is presented below:\(^2\)

\[
\hat{\theta}_j = f_j(t) + \epsilon_j, \quad j = 2, 3, \ldots, p. \tag{8}
\]

Third, we use the angle models of Eq. (8) to predict the angle at time \( T + l \) for different \( j \), as shown in Eq. (9):\(^3\)

\[
\hat{\theta}_j^{T+l} = f_j(T + l), \quad j = 2, 3, \ldots, p. \tag{9}
\]

The predicted value \( \hat{y}_j^{T+l} = (\hat{y}_1^{T+l}, \ldots, \hat{y}_p^{T+l}) \) is computed by using Eq. (6). Since Eq. (6) is derived from the condition \( (r^j)^2 = ||y_j^i||^2 = 1 \), we evidently have

\[
\sum_{j=1}^{p} (\hat{y}_j^{T+l})^2 = 1. \tag{10}
\]

Lastly, we obtain the predicted value of each composition at time \( T + l \) as

\[
x_j^{T+l} = (\hat{y}_j^{T+l})^2, \quad j = 1, 2, \ldots, p. \tag{11}
\]

2.3. Comparison with the logratio transformation

At a given time \( t = 1, 2, \ldots, T \), the well-known logratio transformation proposed by Aitchison (1986) is as follows:

\[
y_j^i = \log \left[ \frac{x_j^i}{x_{j+p}^i} \right], \quad i = 1, 2, \ldots, p - 1. \tag{12}
\]

Aitchison’s (1986) logratio transformation facilitates the modeling functions as \( y_j^i \) is defined on \((0, \infty)\), and as \( y_j^i \) follows the normal distribution when \( x_j^i, \ldots, x_p^i \) follow the additive logistical normal distribution. However, the strict positive value of every composition is a strong assumption. That is,

\[
0 < x_j^i < 1, \quad j = 1, 2, \ldots, p. \tag{13}
\]

Aitchison (1986) provides three ways out of this modeling dilemma: (a) an amalgamation of the finer parts to remove the zero components, (b) replacement of the zeros with very small values, and (c) treatment of the zero observations as outliers. However, considering zero observations as outliers would not be possible if one has many such zero observations in the data set. The zero-replacement procedure would be appropriate when zero refers to a non-quantifiable proportion. The procedure is also appropriate for data inaccuracy or for rounding off the measurement process. However, a zero value generally means a complete absence, and the logratio transformation will fail in such applications.

In comparison, the assumption about the compositional values in our DRHT approach is as follows:

\[
0 \leq x_j^i < 1, \quad j = 1, 2, \ldots, p. \tag{14}
\]

The DRHT method will fail if and only if one composition equals 1, and all other compositions equal 0. Evidently, our DRHT model compares favorably to Aitchison’s (1986) logratio transformation, as it has a wider application in modeling compositional data.

Our model is also applicable in analyzing categorical data which not only enables the values of the marginal distribution to be summed to one, but also helps the statistical analysis of the independent estimator (Grizzle and Williams, 1972). Furthermore, the trigonometric transformation due to the dimension-reduction approach brought about by a hyperspherical transformation is particularly useful for multinomial distribution calculus.

3. Applications

3.1. Empirical simulation with zero components

To verify the principle and algorithm of the DRHT model, we conduct a numerical simulation with zero components. A set of 4D simulation data, embedded with zero components, is generated by using the non-linear equations as in Eq. (15),

\[
\begin{align*}
x_1^i &= F_i(15) - F_i(0) + \text{normrnd}(0, 0.005) \\
x_2^i &= F_i(30) - F_i(15) + \text{normrnd}(0, 0.005) \\
x_3^i &= F_i(45) - F_i(30) + \text{normrnd}(0, 0.005) \\
x_4^i &= 1 - x_1^i - x_2^i - x_3^i,
\end{align*}
\]

where \( F_i(x) \) is the cumulative distribution function of non-center distribution \( \chi^2(10, 1 + 4t) \), parameter 10 is the degree of freedom of distribution, and parameter \((1 + 4t)\) is the non-center degree changing with time \( t \). Note that in Eq. (15), a term with Gauss noise was added for the composition of \( x_1, \)

\(^2\) Note that other methods can be used to model the data series of the angles.

\(^3\) A limitation of the forecasting results when using Eq. (9) is that Eq. (7) will not be appropriate if and only if one of the \( \hat{\theta}_j^{T+l}, j = 2, 3, \ldots, p \) is equal to 1. In such case, the constraint of the DRHT approach \( x_j^i < 1, j = 1, 2, \ldots, p \) is violated.
We apply the DRHT model to the base period from $t = 1, 2, \ldots, 10$, as well as to the forecasting period in $t = 11, 12$. For $t = 1, 2, \ldots, 10$, we are able to obtain the values of the three angles $h_2^t$, $h_3^t$, and $h_4^t$ after the DRHT transformations of $x_1$, $x_2$, $x_3$, and $x_4$ according to Eq. (4) and Eq. (7) (see Table 2).

Based on the computed angles in Table 2, two third-order polynomial models and one first-order polynomial model are computed for the angles $\theta_2^t$, $\theta_3^t$, and $\theta_4^t$, respectively.

As a result, the predicted values of $\theta_2^t$, $\theta_3^t$, and $\theta_4^t$ at time $t = 11, 12$ can now be calculated as follows:

$$\begin{align*}
\hat{\theta}_2^{11} &= 0.1161 \\
\hat{\theta}_3^{11} &= 0.2451 \\
\hat{\theta}_4^{11} &= 0.6188 \\
\hat{\theta}_2^{12} &= 0.1108 \\
\hat{\theta}_3^{12} &= 0.1187 \\
\hat{\theta}_4^{12} &= 0.5225.
\end{align*}$$

The fitted angles at time $t = 1, 2, \ldots, 10$ and the predicted angles at time $t = 11, 12$ are plotted in Fig. 1.

![Fig. 1. Values of $\theta_2^t$, $\theta_3^t$ and $\theta_4^t$.](image-url)
According to Eq. (6) and Eq. (11), the fitted compositions \( x_1, x_2, x_3, \) and \( x_4 \) at time \( t = 1, 2, \ldots, 10 \), and the predicted compositions at time \( t = 11,12 \) are calculated and reported in Table 3, Table 4, and Fig. 2. Evidently, we predict a zero component which maintains the unit-sum assumption in both the estimation and prediction periods. The forecasting errors (i.e., the difference between the actual values and predicted values) at \( t = 11,12 \) are extremely small (last two columns of Table 4). Thus, the simulation demonstrates the validity of our DRHT approach for compositional data.

### 3.2. Forecasting the trend of the primary, secondary, and tertiary components in the Chinese economy

In this section, we apply our DRHT model to investigate the trend of the three production sectors

![Fig. 2. Observed data and predicted values in the simulation study.](image)
in the Chinese economy from 1990 to 1999, and to predict the sector compositions in years 2000 and 2001. Tertiary activities represent an increasing share in the economy, both in developed and developing countries (e.g., Philippe and Leo, 1999; Davies, 1996). Fisher (1939) was the first to propose the use of the relative proportions of three production sectors, namely, the primary, secondary, and tertiary sectors, to reflect the dynamic process of economic development, as well as the trends of the resource allocations of a region or a country. Davies (1996) showed that people in Malian Sahel shifted to tertiary activities when traditional primary and secondary productions did not guarantee a stable food supply.

In China which has over 10 years of enterprise reform since the 1989 share-holding reform, we observe that the composition of primary, secondary, and tertiary production sectors in the Chinese economy changed drastically. We classify the three production sectors in terms of the proportions of labor absorbed in the three respective sectors (Table 5a).

We then present the computed angles $\theta_2^t$, $\theta_3^t$ for $t = 1990, \ldots, 1999$ (based on Eq. (7)) after a hyperspherical mapping in Table 5b.

Because there are three production sectors, the DRHT procedure produces two second-order polynomial models, and we present the results in Table 5b.

$$\theta_2^t = 0.0007t^2 - 0.0161t + 1.0572$$

$$\theta_3^t = 0.0008t^2 - 0.0214t + 1.1572$$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary industry</td>
<td>0.6010</td>
<td>0.5970</td>
<td>0.5850</td>
<td>0.5640</td>
<td>0.5430</td>
<td>0.5220</td>
<td>0.5050</td>
<td>0.4990</td>
<td>0.4980</td>
<td>0.5010</td>
<td>0.5000</td>
<td>0.5000</td>
<td></td>
</tr>
<tr>
<td>Secondary industry</td>
<td>0.2140</td>
<td>0.2140</td>
<td>0.2170</td>
<td>0.2240</td>
<td>0.2270</td>
<td>0.2300</td>
<td>0.2350</td>
<td>0.2370</td>
<td>0.2350</td>
<td>0.2300</td>
<td>0.2250</td>
<td>0.2230</td>
<td></td>
</tr>
<tr>
<td>Third industry</td>
<td>0.1850</td>
<td>0.1890</td>
<td>0.1980</td>
<td>0.2120</td>
<td>0.2300</td>
<td>0.2480</td>
<td>0.2600</td>
<td>0.2640</td>
<td>0.2670</td>
<td>0.2690</td>
<td>0.2750</td>
<td>0.2770</td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_2^t$</td>
<td>1.0328</td>
<td>1.0313</td>
<td>1.0238</td>
<td>1.0085</td>
<td>0.9968</td>
<td>0.9848</td>
<td>0.9721</td>
<td>0.9674</td>
<td>0.9689</td>
<td>0.9753</td>
</tr>
<tr>
<td>$\theta_3^t$</td>
<td>1.1262</td>
<td>1.1210</td>
<td>1.1097</td>
<td>1.0923</td>
<td>1.0706</td>
<td>1.0495</td>
<td>1.0357</td>
<td>1.0312</td>
<td>1.0278</td>
<td>1.0255</td>
</tr>
</tbody>
</table>

Fig. 3. Computed angles and predicted angles.
To forecast the sector compositions in 2000 and 2001, the predicted angles \( \hat{\theta}_2^{2000}, \hat{\theta}_3^{2000}, \hat{\theta}_2^{2001}, \) and \( \hat{\theta}_3^{2001} \) at \( t = 2000, 2001 \) are computed by using the polynomial models constructed above:

\[
\begin{align*}
\hat{\theta}_2^{2000} &= 0.5065 \\
\hat{\theta}_3^{2000} &= 0.9177 \\
\hat{\theta}_2^{2001} &= 0.5017 \\
\hat{\theta}_3^{2001} &= 0.9063.
\end{align*}
\]

Both the fitted angles at time \( t = 1990, \ldots, 1999 \) and the predicted angles at time \( t = 2000 \) and \( 2001 \) are plotted in Fig. 3.

From the computed angles and the predicted angles \( \hat{\theta}_2^{2000}, \hat{\theta}_3^{2000}, \hat{\theta}_2^{2001}, \) and \( \hat{\theta}_3^{2001} \), we obtain the fitted values of \( x_1, x_2, x_3 \) from 1990 to 1999 and the forecasted values of \( x_1, x_2, x_3 \) in 2000 and 2001 by using Eq. (6) and Eq. (11), respectively (see Table 6a and Fig. 4). Evidence shows that the fitted values capture the pattern well. In comparison with the actual values of the three production sectors, we observe that the forecasting errors (i.e., the fitted values – the actual values) are very small (see Table 6b). Such small forecasting errors confirm the validity of our DRHT model.

### Table 6a
Fitted values and predicted values of the three industries

<table>
<thead>
<tr>
<th>Industry</th>
<th>Year</th>
<th>Fitted</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1  )</td>
<td>0.6134</td>
<td>0.5924</td>
<td>0.5733</td>
</tr>
<tr>
<td>( x_2  )</td>
<td>0.2096</td>
<td>0.2157</td>
<td>0.2209</td>
</tr>
<tr>
<td>( x_3  )</td>
<td>0.1770</td>
<td>0.1918</td>
<td>0.2059</td>
</tr>
<tr>
<td>( \Sigma )</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

### Table 6b
Difference of fitted (predicted) values from actual values

<table>
<thead>
<tr>
<th>Industry</th>
<th>Year</th>
<th>Fitted from actual</th>
<th>Predicted from Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1  )</td>
<td>0.0124</td>
<td>–0.0046</td>
<td>–0.0117</td>
</tr>
<tr>
<td>( x_2  )</td>
<td>–0.0044</td>
<td>0.0017</td>
<td>0.0039</td>
</tr>
<tr>
<td>( x_3  )</td>
<td>–0.008</td>
<td>0.0028</td>
<td>0.0079</td>
</tr>
</tbody>
</table>

Fig. 4. Observed and predicted data of the three industries.
As shown in Fig. 4, there is a fast growth in the tertiary sector ($x_3$) and a fast waning of the primary sector ($x_1$), while that of the secondary sector ($x_2$) stabilized. Furthermore, the model predicts a continuous growth in the tertiary sector. The employment figures show that the tertiary sector absorbs employment faster than the decline in the primary sector. This structural change in employment confirms the observation that the tertiary sector grew rapidly in the Chinese economy in the 1990s. As shown by very small prediction errors, we conclude that the predicted values in 2000 and 2001 also agree with the observed data (last two columns in Tables 6a and 6b).

4. Conclusions

In this paper, we proposed a dimension-reduction approach through hyperspherical transformation (DRHT) for the forecasting of compositional data indexed by time. The DRHT involves a non-linear transformation by mapping a compositional data vector onto a hypersphere, so that the correlated $p$ components are transformed into $p - 1$ independent angles. The non-negativity of the components and the unit-sum constraints are maintained in both the estimation and prediction periods.

A four-dimensional numerical simulation with zero components attests to the usefulness and validity of the new approach. Furthermore, our DRHT model is also successful in uncovering the relationship of the three components of the primary, secondary, and tertiary sectors in the Chinese economy, and in predicting their future sector compositions.

Acknowledgements

The authors wish to acknowledge the financial support of the National Natural Science Foundation of China (NSFC), grant number 70371007 and 70125003, and a direct grant from The Chinese University of Hong Kong.

Appendix A. Dimensionality reduction through hyperspherical transformation

The DRHT involves a non-linear transformation by mapping a compositional data vector onto a hypersphere in order to reduce its dimension from $p$ to $p - 1$. We illustrate the DRHT through a set of 3D compositional data vector.

$$X' = \{ (x'_1, x'_2, x'_3) \in \mathbb{R}^3 \mid \sum_{j=1}^{3} x'_j = 1, 0 \leq x'_j < 1 \},$$

$t = 1, 2, \ldots, T$.

We use a square root transformation to reduce the dimension of these vectors. Obviously, the degree of freedom of the above vector equals to 2 due to the unit-sum constraint.

$$y' = \sqrt{x'_j}, \quad j = 1, 2, 3 \quad t = 1, 2, \ldots, T.$$  

Since $y' = (y'_1, y'_2, y'_3)$ is a non-negative vector, it appears in the first quadrant of the Cartesian coordinate system, and

$$||y'||^2 = \sum_{j=1}^{3} (y'_j)^2 = 1.$$  

Next, we map the vector $y' = (y'_1, y'_2, y'_3)' \in \mathbb{R}^3(t = 1, 2, \ldots, T)$ from the Cartesian coordinate system onto a spherical coordinate system $(r', \theta'_2, \theta'_3)' \in \Theta^3$. Considering the condition $(r')^2 = ||y'||^2 \equiv 1$, the mapping equations would be as follows (see also Fig. A1):

$$\begin{align*}
    y'_1 &= \sin \theta'_2 \sin \theta'_3 \\
    y'_2 &= \cos \theta'_2 \sin \theta'_3 \\
    y'_3 &= \cos \theta'_3,
\end{align*}$$

where $0 < \theta'_3 \leq \pi/2, j = 2, 3$.

Evidently, all the vectors of $y' = (y'_1, y'_2, y'_3)'$, $t = 1, 2, \ldots, T$ are on the surface of a 1/8 sphere with radius = 1. The three linearly related variables $y'_1, y'_2, y'_3$ changed into two independent angles $\theta'_2, \theta'_3$. Thus, we reduce the dimension of the vector from three to two.

Fig. A1. Mapping vector from the 3D coordinate to the sphere.
References

Pearson, K., 1897. Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs. Proceedings of the Royal Society of London 60, 489–498.