AN ALGEBRAIC METHOD TO COMPUTE ISOMETRIC LOGRATIO TRANSFORMATION AND BACK TRANSFORMATION OF COMPOSITIONAL DATA

Eusebi Jarauta-Bragulat[1], **Pablo Buenestado-Caballero**[2], **Carme Hervada-Sala**[3]

[1]Department of Applied Mathematics III, ETSECCPB/EUETIB.
Universitat Politècnica de Catalunya (UPC). Barcelona, Spain. Eusebi.Jarauta@upc.es.
[2]Department of Applied Mathematics III, EUETIB.
Universitat Politècnica de Catalunya (UPC). Barcelona, Spain. Pablo.Buenestado@upc.es.
[3]Department of Physics and Nuclear Engineering, EUETIT.
Universitat Politècnica de Catalunya (UPC). Terrassa, Spain. Carme.Hervada@upc.es.

Abstract

One of the most important aspects associated with transformations of compositional data, consists of establishing the analytical expression of the data transformation and, specially, the inverse transformation; in that case, once data of the simplex have been transformed in elements of the ordinary Euclidian space to proceed to its treatment (for example, kriging), apply to them the back transformation that allows to obtain again the corresponding elements in the simplex. Compositional data transformations more commonly applied up to now, are the additive logratio transformation (ALR) and the centered logratio transformation (CLR), both defined by J. Aitchison. Recently it was established by J. J. Egozcue and others (2003), a new compositional data transformation, named isometric logratio (ILR), with the purpose of having an orthonormal basis of reference in the simplex. The proposed expression for the calculation of the data transformation and the inverse transformation are really complex and they are based on operations defined in the simplex: inner product of Aitchison, linear closure and power transformation. Furthermore, those expressions may present occasionally error problems by round and truncation. In this paper, a simple method for the calculation of the ILR data transformation and ILR inverse transformation is proposed, based on algebraic methods and it could be easily implemented into a spread-sheet.

Introduction

In the area of the Earth sciences, when we study *n* compositional characteristics, $V_1, V_2, \ldots, V_n$, then the data set is a matrix $D \in \mathfrak{M}_{\mathbb{R}}(p,n)$, that is, a set of *p* cases (rows), such that to each case correspond to its *n* compositional values (columns). In order to complete adequately the matrix data, we define the residual variable V_R, through the relationship: $V_R = K_{CL} - (V_1 + V_2 + \cdots + V_n)$, where K_{CL} is the closure constant, that is, $K_{CL} = 1$ if the data are proportions respect to unit and $K_{CL} = 100$ if it is considered percentages, etc. Hence, considering the studied *n* variables and the residual variable, in this case each datum, $x = (x_1, x_2, \ldots, x_n, x_R)$, is a compositional vector of the simplex S_{n+1}, that it is a vector space on $\mathbb{R}$ of dimension n, with the $\oplus$ (closure) and $\otimes$ (power by scalar) operations.

Consider the hyper plane *H* (vector subspace of dimension *n* in the $\mathbb{R}^{n+1}$ ordinary Euclidean space) defined by the equation $\alpha_1 + \alpha_2 + \cdots + \alpha_{n+1} = 0$. A (natural) base of *H* is formed by the n vectors:

$$\vec{w}_1 = (1,-1,0,...,0),\ \vec{w}_2 = (0,1,-1,...,0),\ ...,\ \vec{w}_n = (0,...,0,1,-1) \tag{1.1}$$

Applying to this base the method of Gram-Schmidt orthogonalization, we obtain the orthogonal base:

$$\vec{v}_1 = (1,-1,0,...,0),\ \vec{v}_2 = \left(\frac{1}{2},\frac{1}{2},-1,...,0\right),\ ...,\ \vec{v}_n = \left(\frac{1}{n},...,\frac{1}{n},-1\right) \tag{1.2}$$

Dividing each vector for its norm, we obtain the orthonormal base:

$$\vec{u}_1 = \sqrt{\frac{1}{2}}\left(1,-1,0,\ldots,0\right),\ \vec{u}_2 = \sqrt{\frac{2}{3}}\left(\frac{1}{2},\frac{1}{2},-1,\ldots,0\right),\ \ldots,\ \vec{u}_n = \sqrt{\frac{n}{n+1}}\left(\frac{1}{n},\ldots,\frac{1}{n},-1\right) \tag{1.3}$$

In these conditions, we define the isometric logratio transformation, denoted by ILR, as the mapping $\varphi : S_{n+1} \to H$ defined by:

$$\vec{y} = \varphi(\vec{x}) = \sum_{k=1}^{n} < \ln(\vec{x}), \vec{u}_k > \vec{u}_k,\ \ \vec{x} \in S_{n+1} \tag{1.4}$$

where $<\ ,\ >$ indicates the ordinary Euclidean inner product; it is a simple exercise to verify that the φ mapping is an isomorphism of vector spaces. For the calculation of the inverse transformation, Egozcue and others (2003) establish the following methodology: if CL indicates the closure respect to the constant K_{CL} and is designated through:

$$\vec{e}_k = CL(\exp(\vec{u}_k)) = (e_k^1, e_k^2, \ldots, e_k^{n+1}) \in S_{n+1},\ \ k = 1,2,\ldots,n \tag{1.5}$$

then:

$$\varphi^{-1}(\vec{y}) = \varphi^{-1}(y_1,\ldots,y_n) = CL(\ldots,\ (e_1^j)^{y_1}\cdot(e_2^j)^{y_2}\cdot\cdots\cdot(e_n^j)^{y_n},\ldots\ ) \in S_{n+1},\ j = 1,2,\ldots,n+1 \tag{1.6}$$

Aside from the inherent operative difficulty to its definition, it has been proven by Jarauta-Bragulat and others (2003) that the expressions (1.4) and (1.6) present computacional problems for truncation errors. For this reason and with the objective to obtain a simplest and operative expression, in this article is proposed a methodology based on the matrix algebra.

Methodology

Developing the equation (1.4), it is obtained that the *k*th component of the vector $\vec{y} = \varphi(\vec{x})$ in the orthonormal base (1.3), is given by:

$$y_k = \sqrt{\frac{k}{k+1}}\ \ln\frac{g(x_1, x_2, \ldots, x_k)}{x_{k+1}},\ \ k = 1,2,\ldots,n \tag{2.1}$$

where $g(x_1, x_2, \ldots, x_k)$ indicates the geometric mean of the corresponding elements. Developing in equation (2.1) the expression of the geometric mean and taking into account the properties of the natural logarithm, the result to be the following equations, that allows us to calculate the ILR transformation as linear combination of the components of the data vector:

$$\sqrt{\frac{k+1}{k}}y_k = \frac{1}{k}\ln x_1 + \frac{1}{k}\ln x_2 + \cdots + \frac{1}{k}\ln x_k - \ln x_{k+1},\ \ k = 1,2,\ldots,n \tag{2.2}$$

This allows us to write the equations of ILR transformation by a simple matricial equation, through:

$$\begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 1/2 & 1/2 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 1/3 & 1/3 & 1/3 & -1 & 0 & \cdots & 0 & 0 \\ \cdots & & \cdots & & & & & \\ 1/k & 1/k & \cdots & 1/k & 1/k & -1 & \cdots & 0 \\ \cdots & & \cdots & & & & & \\ 1/n & 1/n & & \cdots & & & 1/n & -1 \end{pmatrix} \begin{pmatrix} \ln x_1 \\ \ln x_2 \\ \\ \vdots \\ \\ \ln x_n \\ \ln x_R \end{pmatrix} = \begin{pmatrix} \sqrt{2}y_1 \\ \sqrt{\frac{3}{2}}y_2 \\ \\ \vdots \\ \\ \sqrt{\frac{n+1}{n}}y_n \end{pmatrix} \iff M^* X = Y \tag{2.3}$$

where $M^* \in \mathfrak{M}_{\mathbb{R}}(n,n+1)$, $X \in \mathfrak{M}_{\mathbb{R}}(n+1,1)$ and $Y \in \mathfrak{M}_{\mathbb{R}}(n,1)$. To obtain the compositional variables from ILR transformed values, the procedure would be immediate if the matrix M^* would be regular, since would be enough to multiply both members by its inverse matrix. Nevertheless, this matrix is not even squared. We propose the following methodology to apply this idea: define a new matrix M by concatenation for rows of M^* and a new row matrix, so that the resulting matrix will be regular, to introduce a new variable and to obtain finally a matricial equation from which one could be obtained the wished solution. Below, we go to develop and to justify the methodology proposed.

If $\vec{e}_{n+1} = (0,0,...,0,1)$ is the (n+1)-th vector of the natural base of $\mathbb{R}^{n+1}$, we can obtain a new matrix $M \in \mathfrak{M}_{\mathbb{R}}(n+1)$ by concatenation of the rows from M^* with that vector and also a new matrix $Y_Z \in \mathfrak{M}_{\mathbb{R}}(n+1,1)$ by concatenation of Y with a new variable $Z = \ln x_R$; that is:

$$M = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 1/2 & 1/2 & -1 & 0 & 0 & \cdots & \cdots & 0 \\ \cdots & & \cdots & & & & & \\ 1/n & 1/n & \cdots & & \cdots & 1/n & 1/n & -1 \\ 0 & 0 & 0 & & \cdots & & 0 & 1 \end{pmatrix} = \begin{pmatrix} c_1 & c_2 & \cdots & c_{n+1} \end{pmatrix}; \quad Y_Z = \begin{pmatrix} \sqrt{2}y_1 \\ \sqrt{\frac{3}{2}}y_2 \\ \vdots \\ \sqrt{\frac{n+1}{n}}y_n \\ Z \end{pmatrix} \tag{2.4}$$

In this way, it can be written the matrix equation $MX = Y_Z$, that gives cause for a system of n+1 linear equations whose first n equations are the same with those which are obtained from the equation (2.3). Our purpose is now to demonstrate that this new matricial equation allows us to obtain the wanted solution.

Proposition 1. - The matrix M is regular anyone that it will be $n \in \mathbb{N}$.

Demonstration: We will prove that $\det M \neq 0$, for all $n \in \mathbb{N}$; in particular, we will show that $\det M = 1$, for all number $n \in \mathbb{N}$. It will be applied that the determinant of a triangular squared matrix by blocks equals to the product of the determinant of the blocks of the diagonal, it supposed squared matrices, and also that the determinant of a triangular squared matrix, upper and lower, is the product of the elements of the main diagonal. For this, it will be divided M in a way convenient to be triangular by blocks and then it is triangularized one of the blocks through elemental transformations by rows. We proceed by induction on *n*.

1) Case $n = 2$. In this case we have:

$$M = \begin{pmatrix} 1 & -1 & 0 \\ 1/2 & 1/2 & -1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} A_2 & B \\ 0 & 1 \end{pmatrix}; \quad \det M = (\det A_2).1 = \det A_2 = 1 \tag{2.5}$$

2) Case $n = 3$. It is possible to show without difficulty that $\det M = 1$.
3) Induction hypothesis. We suppose that to $n-1$ is satisfied:

$$M = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 1/2 & 1/2 & -1 & 0 & \cdots & 0 \\ \cdots & & \cdots & & & \\ 1/(n-1) & \cdots & & & 1/(n-1) & -1 \\ 0 & 0 & \cdots & & 0 & 1 \end{pmatrix} = \begin{pmatrix} A_{n-1} & B \\ 0 & 1 \end{pmatrix}; \quad \det M = \det A_{n-1} = 1 \tag{2.6}$$

4) General case. We decompose the matrix M in blocs: $M = \begin{pmatrix} A_n & B \\ 0 & 1 \end{pmatrix}$, being verified then that $\det M = \det A_n$. Then, in this case it is satisfied:

$$\det A = \det \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 1/2 & 1/2 & -1 & 0 & \cdots & 0 \\ \cdots & & \cdots & & & \\ 1/(n-1) & \cdots & & & 1/(n-1) & -1 \\ 1/n & 1/n & \cdots & & 1/n & 1/n \end{pmatrix} =$$
$$= \det \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 1/2 & 1/2 & -1 & 0 & \cdots & 0 \\ \cdots & & \cdots & & & \\ n/(n-1) & \cdots & & & n/(n-1) & 0 \\ 1/n & 1/n & \cdots & & 1/n & 1/n \end{pmatrix} = \frac{1}{n} \cdot n \cdot \det A_{n-1} = 1 \tag{2.7}$$

and that completes the demonstration. ■

As consequence, the matrix M is regular and admits inverse, M^{-1}, that allows us to obtain the solution that were sought, that is:

$$MX = Y_Z \quad \Leftrightarrow \quad X = M^{-1} Y_Z \tag{2.8}$$

To develop thoroughly this equation, it must be obtained the general expression from the matrix M^{-1}, and this is what we do in the following proposition.

Proposition 2. - The general expression of the matrix M^{-1}, inverse of $M \in \mathfrak{M}_{\mathbb{R}}(n+1)$, is:

$$M^{-1} = \begin{pmatrix} 1/2 & 1/3 & \cdots & 1/n & 1 & 1 \\ -1/2 & 1/3 & \cdots & 1/n & 1 & 1 \\ 0 & -2/3 & \cdots & 1/n & 1 & 1 \\ 0 & 0 & & & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & & -(n-1)/n & 1 & 1 \\ 0 & 0 & & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} r^1 \\ r^2 \\ \vdots \\ r^{n+1} \end{pmatrix} \tag{2.9}$$

Demonstration: It is a simple calculation showing that:

$$c_j r^k = 0 \text{ if } j \neq k; \quad c_j r^k = 1 \text{ if } j = k, \quad j,k = 1,2,\ldots,n+1 \tag{2.10}$$

Algebraic equations of inverse ILR transformation

Once it established the general form of the inverse matrix M^{-1}, we go to obtain the equations from the inverse from the ILR transformation. Developing the matrix product $M^{-1}X$, we obtain the system of linear equations:

$$
\begin{aligned}
\ln x_1 &= \frac{1}{2}\sqrt{2}y_1 + \frac{1}{3}\sqrt{\frac{3}{2}}y_2 + \cdots + \frac{1}{j+1}\sqrt{\frac{j+1}{j}}y_j + \cdots + \frac{1}{n}\sqrt{\frac{n}{n-1}}y_{n-1} + \sqrt{\frac{n+1}{n}}y_n + Z \\
\ln x_k &= -\frac{k-1}{k}\sqrt{\frac{k}{k-1}}y_{k-1} + \frac{1}{k+1}\sqrt{\frac{k+1}{k}}y_k + \cdots + \frac{1}{n}\sqrt{\frac{n}{n-1}}y_{n-1} + \sqrt{\frac{n+1}{n}}y_n + Z, \quad k = 2,3,\dots,n \\
\ln x_R &= Z
\end{aligned}
\tag{3.1}
$$

Applying the exponential function to both members of equations (3.1), it results finally:

$$
\begin{aligned}
x_1 &= \exp\left(\frac{1}{2}\sqrt{2}y_1 + \frac{1}{3}\sqrt{\frac{3}{2}}y_2 + \cdots + \frac{1}{j+1}\sqrt{\frac{j+1}{j}}y_j + \cdots + \frac{1}{n}\sqrt{\frac{n}{n-1}}y_{n-1} + \sqrt{\frac{n+1}{n}}y_n + Z\right), \\
x_k &= \exp\left(-\frac{k-1}{k}\sqrt{\frac{k}{k-1}}y_{k-1} + \frac{1}{k+1}\sqrt{\frac{k+1}{k}}y_k + \cdots + \frac{1}{n}\sqrt{\frac{n}{n-1}}y_{n-1} + \sqrt{\frac{n+1}{n}}y_n + Z\right) \quad (k = 2,3,\dots,n) \\
x_R &= \exp(Z)
\end{aligned}
\tag{3.2}
$$

Usually the available information consists of the ILR coordinates $(y_1, y_2, \dots, y_n)$, and the Z variable is unknown. In such a case, we observe that in equation (3.2) the first n equations can be written as:

$$
x_k = \exp(S_k + Z) = \exp(S_k)\exp(Z) \iff \exp(S_k) = \frac{1}{\exp(Z)}x_k \quad (k = 1,2,\dots,n)
$$

Then, if we say $\alpha = \sum_{k=1}^{n}\exp(S_k)$, we obtain:

$$
\alpha = \frac{1}{\exp(Z)}\sum_{k=1}^{n}x_k = \frac{1}{x_R}(1 - x_R) \Rightarrow x_R = \frac{1}{1+\alpha}
$$

and we can get the expression of the Z variable:

$$
Z = \ln x_R = \ln\frac{1}{1+\alpha} = -\ln(1+\alpha) \tag{3.3}
$$

Finally, once Z variable is obtained, equations (3.2) can be applied to obtain the remaining compositional coordinates $x_1, x_2, \dots, x_n$.

An example

To illustrate the proposed methodology, there is a simple example with four compositional variables V_1, V_2, V_3, V_R; in Table 1 you can see the data and its ILR transformations Y_1, Y_2, Y_3. The equations to compute the back transformed values are:

$$
\begin{aligned}
&x_1 = \exp\left(\sqrt{\frac{1}{2}}y_1 + \sqrt{\frac{1}{6}}y_2 + \sqrt{\frac{4}{3}}y_3 + Z\right); \quad x_2 = \exp\left(-\sqrt{\frac{1}{2}}y_1 + \sqrt{\frac{1}{6}}y_2 + \sqrt{\frac{4}{3}}y_3 + Z\right); \\
&x_3 = \exp\left(-\sqrt{\frac{2}{3}}y_2 + \sqrt{\frac{4}{3}}y_3 + Z\right)
\end{aligned}
\tag{4.1}
$$

From ILR transformed values, we can compute $\exp(S_k),\ k = 1,2,3$ and from those values, parameter α can be also computed, as its sum. So, we can compute the Z variable by applying equation (3.3) and finally, back transformed values can be computed applying equations (3.2). All those computations are shown in Table 2.

Table 1. Data matrix (V1,V2,V3,VR) and ILR transformations (Y1,Y2,Y3)

V1	V2	V3	VR	Y1	Y2	Y3
0,38200	0,10068	0,49650	0,02082	0,94290	-0,75844	2,21048
0,02350	0,43210	0,12340	0,42100	-2,05885	-0,16542	-1,17976
0,01450	0,40742	0,56430	0,01378	-2,35888	-1,62786	2,06383
0,13858	0,24503	0,04547	0,57091	-0,40299	1,14255	-1,38325
0,03238	0,16413	0,21961	0,58388	-1,14771	-0,90040	-1,48351
0,01709	0,28504	0,34309	0,35478	-1,98989	-1,30020	-0,94839
0,23460	0,35737	0,37184	0,03619	-0,29762	-0,20423	1,87313
0,35560	0,09870	0,46602	0,07968	0,90632	-0,74405	1,00345
0,02350	0,30390	0,32140	0,35120	-1,80999	-1,09070	-0,84802
0,45120	0,09980	0,25626	0,19274	1,06684	-0,15405	0,13779

Table 2. Intermediate computations and ILR back transformed values (V*1,V*2,V*3,V*R).

exp(S1)	exp(S2)	exp(S3)	α	Z	V*1	V*2	V*3	V*R
18,34841	4,83594	23,84812	47,03246	-3,872	0,38200	0,10068	0,49650	0,02082
0,05582	1,02637	0,29311	1,37530	-0,865	0,02350	0,43210	0,12340	0,42100
1,05186	29,56274	40,94588	71,56048	-4,284	0,01450	0,40742	0,56430	0,01378
0,24274	0,42920	0,07965	0,75159	-0,561	0,13858	0,24503	0,04547	0,57091
0,05546	0,28110	0,37612	0,71268	-0,538	0,03238	0,16413	0,21961	0,58388
0,04817	0,80344	0,96705	1,81867	-1,036	0,01709	0,28504	0,34309	0,35478
6,48232	9,87468	10,27438	26,63138	-3,319	0,23460	0,35737	0,37184	0,03619
4,46283	1,23869	5,84856	11,55009	-2,530	0,35560	0,09870	0,46602	0,07968
0,06691	0,86534	0,91516	1,84741	-1,046	0,02350	0,30390	0,32140	0,35120
2,34103	0,51781	1,32961	4,18844	-1,646	0,45120	0,09980	0,25626	0,19274

Conclusions

1. The application of isometric logratio transformation (ILR) and its inverse to compositional data can become difficult in some cases, as well as to show numerical error problems of truncation.
2. The methodology proposed in this article, based on standard algebraic methods, results notably simpler and it is easily implemented into a spread-sheet.
3. The methodology proposed in this article operates correctly, according to what has been verified.

References

Egozcue, J.J., V. Pawlowsky-Glahn, G. Mateu-Figueras and C. Barceló-Vidal (2003). *Isometric logratio transformations for compositional data analysis.* Mathematical Geology, vol. 35.

Jarauta-Bragulat, E. and C. Hervada-Sala (2003). C*omparison of kriging results of regionalized compositional data using three different data transformations. Case study: bauxites in Hungary.* Submitted for presentation in IAMG'2003.

Jarauta, Eusebi (2001). *Anàlisi Matemàtica d'una variable. Fonaments i aplicacions.* Edicions UPC. 429 pp.

Lang, S. (1990). *Introducción al Álgebra Lineal.* Addison-Wesley Iberoamericana. 265 pp.

Lewis, David W. (1991). *Matrix Theory.* World Scientific. 298 pp.

Acknowledgements.

Authors want to thank specially: the Department of Applied Mathematics III (UPC), the School of Civil Engineering (UPC) and the "Consorci Escola Industrial de Barcelona" (CEIB), for its support in this contribution.