Statistics and Application 统计学与应用, 2026, 15(1), 218-226
Published Online January 2026 in Hans. https://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2026.151021

Hans汉斯

成分数据三种对数比变换的研究

刘若菲，赵沛莹，袁梓佳，关灵欣，黄明佳，尹豫哲

中国矿业大学(北京)理学院，北京

收稿日期：2025年12月19日；录用日期：2026年1月10日；发布日期：2026年1月21日

摘　要

成分数因受“闭合性”约束天然存在于$D-1$维单纯形空间，难以直接应用传统多维统计分析方法。非对称对数比变换(*ALR*)、中心化对数比变换(*CLR*)与等距对数比变换(*ILR*)通过对数比映射将成分数据转化至欧氏空间，是成分数据统计分析的核心工具。本文从变换基、保距性、保角性、矩阵表示、几何意义及应用场景六个维度，系统对比三种变换的差异，明确了各变换的适用边界与优势，为地质学、生态学、医学等多领域的成分数据分析实践提供了理论依据。

关键词

成分数据，对数比变换，艾奇逊距离，单纯形空间

A Study on Three Kinds of Log-Ratio Transformations for Compositional Data

Ruofei Liu, Peiying Zhao, Zijia Yuan, Lingxin Guan, Mingjia Huang, Yuzhe Yin

School of Science, China University of Mining and Technology (Beijing), Beijing

Received: December 19, 2025; accepted: January 10, 2026; published: January 21, 2026

Abstract

Compositional data are naturally constrained by “closure” and exist in a $D-1$ dimensional simplex space, making it difficult to directly apply traditional multivariate statistical methods. Additive log-ratio transformation (*ALR*), centered log-ratio transformation (*CLR*), and isometric log-ratio transformation (*ILR*) convert compositional data to Euclidean space through log-ratio mapping, and they are the core tools for compositional data analysis. This paper systematically compares the differences among the three transformations from six dimensions: transformation basis, distance preservation, angle preservation, matrix representation, geometric meaning, and application scenarios. The research clarifies the applicable boundaries and advantages of each transformation, providing a theoretical basis for the practice of compositional data analysis in multiple fields such as geology,

文章引用: 刘若菲, 赵沛莹, 袁梓佳, 关灵欣, 黄明佳, 尹豫哲. 成分数据三种对数比变换的研究[J]. 统计学与应用, 2026, 15(1): 218-226. DOI: 10.12677/sa.2026.151021

ecology, and medicine.

Keywords

Compositional Data, Log-Ratio Transformation, Aitchison Distance, Simplex Space

1. 引言

成分数据是描述整体由部分构成的特殊数据形态——它以比例、百分比或浓度等形式存在，核心特征是所有分量的总和恒为常数(如 1、100%或 10^6 ppm)，且各分量均为非负值。

从古人通过“五谷配比”总结农耕经验，到现代科学家解析母乳中 2000 余种营养成分的协同作用，从考古学家通过陶瓷釉层元素占比溯源窑口，到经济学家分析家庭消费结构的变迁规律，成分数据早已渗透到自然科学与社会科学的每一个角落。

然而，这种“定和约束”的特性，使其与传统数据存在本质区别：将常规统计方法直接应用于成分数据，会产生虚假相关、负偏差等悖论，例如生态研究中某物种占比上升必然导致其他物种占比下降，导致各成分协方差阵出现无意义的负值。

直至 1986 年，英国统计学家艾奇逊(J. Aitchison)出版《成分数据的统计分析》[1]，构建了成分数据的艾奇逊空间，标志着成分数据分析成为独立学科。

在艾奇逊空间中，设成分向量 $x=(x_1,\cdots,x_D)^{\mathrm{T}}, y=(y_1,\cdots,y_D)^{\mathrm{T}}\in S^{D-1}$，则扰动运算定义为 $x\oplus y=\left(\frac{x_1y_1}{\sum_{i=1}^{D}x_iy_i},\cdots,\frac{x_Dy_D}{\sum_{i=1}^{D}x_iy_i}\right)^{\mathrm{T}}$；幂运算定义为 $x\odot\lambda=\left(\frac{x_1^{\lambda}}{\sum_{i=1}^{D}x_i^{\lambda}},\cdots,\frac{x_D^{\lambda}}{\sum_{i=1}^{D}x_i^{\lambda}}\right)^{\mathrm{T}}$；内积定义为 $\langle x,y\rangle_{\alpha}=\sum_{i=1}^{D}\ln\frac{x_i}{\sqrt[D]{x_1\cdots x_D}}\ln\frac{y_i}{\sqrt[D]{y_1\cdots y_D}}$；距离定义为：$d_{\alpha}\langle x,y\rangle=\sqrt{\frac{1}{2}\sum_{i=1}^{D}\sum_{j=1}^{D}\left(\ln\frac{x_i}{x_j}-\ln\frac{y_i}{y_j}\right)^2}$。

针对成分数据因“定和约束”导致的传统统计方法失效问题，艾奇逊在 1986 年首次提出中心对数比变换(*CLR*)，将单纯形上的成分数据映射至欧氏空间，还对已有的非对称对数比变换(*ALR*)进行了系统梳理与性质分析，明确其作为 *CLR* 特例的数学定位；2003 年，Egozcue 等进一步优化该体系，提出等距对数比变换(*ILR*) [2]，通过正交基构造解决了 *ALR* 的非等距性与 *CLR* 多重共线性问题，完善了成分向量的向量空间结构。

这三种变换分别为 $ALR(x)=\left(\ln\frac{x_1}{x_D},\cdots,\ln\frac{x_{D-1}}{x_D}\right)^{\mathrm{T}}$；$CLR(x)=\left(\ln\frac{x_1}{\sqrt[D]{x_1\cdots x_D}},\cdots,\ln\frac{x_D}{\sqrt[D]{x_1\cdots x_D}}\right)^{\mathrm{T}}$；$ILR(x)=\sqrt{\frac{i}{i+1}}\left(\ln\frac{x_1}{x_2},\ln\frac{\sqrt{x_1x_2}}{x_3},\cdots,\ln\frac{\sqrt[D-1]{x_1x_2\cdots x_{D-1}}}{x_D}\right)^{\mathrm{T}}$。

三种变换的转换基及逆变换是不同的。*ALR* 变换的转换基为 $\left(e_1=c(e,1,\cdots,1,1)^{\mathrm{T}},\cdots,e_{D-1}=c(1,1,\cdots,e,1)^{\mathrm{T}}\right)^{\mathrm{T}}$，相应逆变换为 $x_i=\frac{e^{y_i}}{\sum_{j=1}^{D-1}e^{y_i}+1}, i=1,\cdots,D-1; x_D=\frac{1}{\sum_{j=1}^{D-1}e^{y_i}+1}$；*CLR* 变换的转换基：

$\left(e_1=c\left(e,\cdots,1\right)^{\mathrm{T}},\cdots,e_D=c\left(1,\cdots,e\right)^{\mathrm{T}}\right)^{\mathrm{T}}$，相应逆变换为 $x_i=\dfrac{e^{y_i}}{\sum_{j=1}^{D}e^{y_i}},i=1,\cdots,D-1;x_D=\dfrac{1}{\sum_{j=1}^{D}e^{y_i}}$；*ILR* 变换的转换基 $\left(e_1,\cdots,e_D\right)^{\mathrm{T}}$ 为一组正交基，相应逆变换为 $x=\oplus_{i=1}^{D-1}\left(y_i\cdot e_i\right)$。

艾奇逊空间作为成分数据的数学理论基础(含扰动运算、幂运算、艾奇逊内积与距离定义)，厘清了 *ALR*，*CLR*，*ILR* 三种变换的提出脉络、核心公式与本质目标——通过从单纯形空间到欧氏空间的映射，彻底打破成分数据的分析桎梏。

接下来就这三种变换详细分析它们的异同。

2. 三种对数比变换的差异性

(一) 等距性

所谓等距性指的是在单纯形与欧氏空间的转换过程中，距离度量属性得以保留，即欧式空间中两点的距离与两点经过相应对数比逆变换后的艾奇逊空间中两点的距离相等[1]。

定理 1：对于任意成分向量 $\boldsymbol{x}=\left(x_1,\cdots,x_D\right)^{\mathrm{T}},\boldsymbol{y}=\left(y_1,\cdots,y_D\right)^{\mathrm{T}}\in S^{D-1}$，有

$$d_a\left(x,y\right)=\left\|CLR\left(x\right)-CLR\left(y\right)\right\|=\left\|ILR\left(x\right)-ILR\left(y\right)\right\|\neq\left\|ALR\left(x\right)-ALR\left(y\right)\right\|$$

证明：令 $\ln\dfrac{x_j}{x_D}=y_{1j}$，$\ln\dfrac{y_j}{y_D}=y_{2j}$，则

$$d_\alpha^2\left(x,y\right)=\left\|x\oplus\left(-1\right)\otimes y\right\|^2=\left\langle x\oplus\left(-1\right)\otimes y,x\oplus\left(-1\right)\otimes y\right\rangle_\alpha=\left\langle x,x\right\rangle_\alpha-2\left\langle x,y\right\rangle_\alpha+\left\langle y,y\right\rangle_\alpha$$

$$\text{又}\left\langle x,x\right\rangle_\alpha=\sum_{i=1}^{D}\left(\ln\frac{x_i}{g_m\left(x\right)}\right)^2=\sum_{i=1}^{D}\left(\ln\frac{x_D\mathrm{e}^{y_{1i}}}{x_D\left(\mathrm{e}^{\sum_{j=1}^{D}y_{1j}}\right)^{1/D}}\right)^2=\sum_{i=1}^{D}\left(y_{1i}-\frac{1}{D}\sum_{j=1}^{D}y_{1j}\right)^2$$

同理：

$$\left\langle y,y\right\rangle_\alpha=\sum_{i=1}^{D}\left[y_{2i}-\frac{1}{D}\sum_{j=1}^{D}y_{2j}\right]^2;\left\langle x,y\right\rangle_\alpha=\sum_{i=1}^{D}\left[y_{1i}-\frac{1}{D}\sum_{j=1}^{D}y_{1i}\right]\left[y_{2i}-\frac{1}{D}\sum_{j=1}^{D}y_{2j}\right]$$

整理可得

$$d_\alpha^2\left(x,y\right)=\sum_{i=1}^{D}\left(y_{1i}-y_{2i}\right)^2+\frac{1}{D}\left(\sum_{j=1}^{D}y_{1j}-\sum_{j=1}^{D}y_{2j}\right)^2\neq\sum_{i=1}^{D-1}\left(y_{1i}-y_{2i}\right)^2=\left\|ALR\left(x\right)-ALR\left(y\right)\right\|$$

CLR 等距证明与 *ALR* 的类似，这里不再展现。

对于 *ILR*，令 $x=\oplus_{i=1}^{D-1}y_{1i}\otimes e_i$，$y=\oplus_{j=1}^{D-1}y_{2i}\otimes e_i$，其中 $e_i\left(i=1,2,\cdots,D-1\right)$ 为艾奇逊标准正交基，

$\left\langle x,x\right\rangle_\alpha=\left\langle\oplus_{i=1}^{D-1}y_{1i}\otimes e_i,\oplus_{j=1}^{D-1}y_{1j}\otimes e_j\right\rangle_\alpha=\sum_{i=1}^{D-1}\sum_{j=1}^{D-1}y_{1i}y_{1j}\left\langle e_i,e_j\right\rangle_a=\sum_{i=1}^{D-1}y_{1i}^2$；

$\left\langle x,y\right\rangle_a=\left\langle\oplus_{i=1}^{D-1}y_{1i}\otimes e_i,\oplus_{j=1}^{D-1}y_{2j}\otimes e_j\right\rangle_a=\sum_{i=1}^{D-1}\sum_{j=1}^{D-1}y_{1i}y_{2j}\left\langle e_i,e_j\right\rangle_a=\sum_{i=1}^{D-1}y_{1i}y_{2i}$，所以

$d_\alpha^2\left(x,y\right)=\sum_{i=1}^{D-1}y_{1i}^2-\sum_{j=1}^{D}2y_{1i}y_{2i}+\sum_{i=1}^{D-1}y_{2i}^2=\sum_{i=1}^{D-1}\left(y_{1i}-y_{2i}\right)^2=\left\|ILR\left(x\right)-ILR\left(y\right)\right\|$。

ALR 不具等距性，因其丢弃了参考成分与其他成分间的部分距离信息；*CLR* 变换具有等距性，但是 *CLR* 系数有零和约束；*ILR* 变换亦具有等距性，使其广泛应用于聚类、主成分分析等依赖距离度量的统计分析中。

(二) 保角性

所谓保角性指的是两成分向量在艾奇逊空间的夹角与变换后相应向量在欧氏空间的夹角保持一致[1]。对于成分向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$，艾奇逊夹角定义为 $\langle \boldsymbol{x},\boldsymbol{y}\rangle = \dfrac{\langle \boldsymbol{x},\boldsymbol{y}\rangle_\alpha}{d_\alpha\langle \boldsymbol{x},\boldsymbol{y}\rangle}$。

定理 2：对于任意成分向量 $\boldsymbol{x}=(x_1,\cdots,x_D)^{\mathrm{T}}, \boldsymbol{y}=(y_1,\cdots,y_D)^{\mathrm{T}} \in S^{D-1}$，有：

$$\langle \boldsymbol{x},\boldsymbol{y}\rangle = \langle CLR(\boldsymbol{x}),CLR(\boldsymbol{y})\rangle = \langle ILR(\boldsymbol{x}),ILR(\boldsymbol{y})\rangle \neq \langle ALR(\boldsymbol{x}),ALR(\boldsymbol{y})\rangle$$

证明：

$$\begin{aligned}
&ALR(\boldsymbol{x})\cdot ALR(\boldsymbol{y}) - \langle \boldsymbol{x},\boldsymbol{y}\rangle_\alpha \\
&= \Big[\sum \ln x_i \ln y_i - \ln x_D \sum \ln y_i - \ln y_D \sum \ln x_i + D\ln x_D \ln y_D\Big] \\
&\quad -\Big[\sum \ln x_i \ln y_i - \ln g(x)\sum \ln y_i - \ln g(y)\sum \ln x_i + D\ln g(x)\ln g(y)\Big] \\
&= D\Big[-\ln x_D \ln g(y) + \ln g(x)\ln g(y) - \ln y_D \ln g(x) + \ln x_D \ln y_D\Big] \\
&= D\ln\frac{x_D}{g(x)}\ln\frac{y_D}{g(y)}
\end{aligned}$$

即 $ALR(\boldsymbol{x})\cdot ALR(\boldsymbol{y}) = \langle \boldsymbol{x},\boldsymbol{y}\rangle_\alpha + D\ln\dfrac{x_D}{g(x)}\ln\dfrac{y_D}{g(y)}$，

故 $\langle \boldsymbol{x},\boldsymbol{y}\rangle \neq \langle ALR(\boldsymbol{x}),ALR(\boldsymbol{y})\rangle$。

又 $CLR(\boldsymbol{x})\cdot CLR(\boldsymbol{y}) = \sum_{i=1}^{D}\ln\dfrac{x_i}{g(x)}\ln\dfrac{y_i}{g(y)} = \langle \boldsymbol{x},\boldsymbol{y}\rangle_\alpha$，且 $d_\alpha(\boldsymbol{x},\boldsymbol{y}) = \|CLR(\boldsymbol{x}) - CLR(\boldsymbol{y})\|$，故 $\langle \boldsymbol{x},\boldsymbol{y}\rangle = \langle CLR(\boldsymbol{x}),CLR(\boldsymbol{y})\rangle$。

由定理 1 可知 $ILR(\boldsymbol{x})\cdot ILR(\boldsymbol{y}) = \langle \boldsymbol{x},\boldsymbol{y}\rangle_\alpha$，欧氏模长：$\|ILR(\boldsymbol{x})\| = \sqrt{ILR(\boldsymbol{x})\cdot ILR(\boldsymbol{x})} = \sqrt{\langle \boldsymbol{x},\boldsymbol{x}\rangle_\alpha} = \|\boldsymbol{x}\|_\alpha$，同理 $\|ILR(\boldsymbol{y})\| = \|\boldsymbol{y}\|_\alpha$，故 $\langle \boldsymbol{x},\boldsymbol{y}\rangle = \dfrac{\langle \boldsymbol{x},\boldsymbol{y}\rangle_\alpha}{d_\alpha\langle \boldsymbol{x},\boldsymbol{y}\rangle} = \dfrac{\langle \boldsymbol{x},\boldsymbol{y}\rangle_\alpha}{\|\boldsymbol{x}\|_\alpha\|\boldsymbol{y}\|_\alpha} = \dfrac{ILR(\boldsymbol{x})\cdot ILR(\boldsymbol{y})}{\|ILR(\boldsymbol{x})\|\|ILR(\boldsymbol{y})\|} = \langle ILR(\boldsymbol{x}),ILR(\boldsymbol{y})\rangle$，即 $\langle \boldsymbol{x},\boldsymbol{y}\rangle = \langle ALR(\boldsymbol{x}),ALR(\boldsymbol{y})\rangle$。

ILR 变换是艾奇逊空间到欧氏空间的等距同构，具有内积和距离完全保持的性质，同时 *ILR* 变换也是保角变换，这是 *ILR* 变换得以在成分数据统计分析中广泛应用的原因之一。

(三) 矩阵表示及转换

对于成分向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_{D-1}) \in \boldsymbol{S}^{D-1}$，三种对数比变换的矩阵表示各不相同。

ALR 变换矩阵 $\boldsymbol{F}_{D*(D-1)} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ -1 & -1 & \cdots & -1 & -1 \end{pmatrix}$，满足 $ALR(\boldsymbol{x})_{1*(D-1)} = \ln(\boldsymbol{x}_{1*D})\boldsymbol{F}_{D*(D-1)}$。结构颇具规律性，即：$\boldsymbol{F}_{D\times(D-1)} = \begin{pmatrix} \boldsymbol{E}_{D-1} \\ -\boldsymbol{1}_{1\times(D-1)} \end{pmatrix}$，其中 $\boldsymbol{1}_{1\times(D-1)}$ 是元素全为 1 的行向量。

性质 1：1) 列线性无关性。变换矩阵 $\boldsymbol{F}$ 的 $D-1$ 个列向量线性无关，保证了变换后的数据无信息丢失，保证了变换的可逆性。

2) 左零空间特性。左零空间为全 1 行向量的张成空间 $span\{\boldsymbol{1}_{1\times D}\}$，这一特性对应成分数据的归一化约束。

CLR 变换矩阵 $\boldsymbol{B}_{D*D}=\begin{pmatrix}\frac{D-1}{D} & -\frac{1}{D} & \cdots & -\frac{1}{D} & -\frac{1}{D}\\ -\frac{1}{D} & \frac{D-1}{D} & \cdots & -\frac{1}{D} & -\frac{1}{D}\\ \vdots & \vdots & \cdots & \vdots & \vdots\\ -\frac{1}{D} & -\frac{1}{D} & \cdots & \frac{D-1}{D} & -\frac{1}{D}\\ -\frac{1}{D} & -\frac{1}{D} & \cdots & -\frac{1}{D} & \frac{D-1}{D}\end{pmatrix}$，满足 $CLR(\boldsymbol{x})_{1*D}=\ln(\boldsymbol{x}_{1*D})\boldsymbol{B}_{D*D}$。结构颇具规律性，即：$\boldsymbol{B}_{D\times D}=\boldsymbol{E}_D-\frac{1}{D}\boldsymbol{1}_{D\times D}$。

性质 2：1) 正交性。矩阵的行和与列和均为零，是对称阵，行向量均与全 1 向量正交。

2) 幂等性。满足 $\boldsymbol{B}^2=\boldsymbol{B}$，即 CLR 变换是投影变换，对数向量经 $\boldsymbol{B}$ 投影一次后，再投影一次的结果与第一次完全相同。

ILR 变换矩阵为 $\boldsymbol{\Psi}=(\psi_{ij})_{(D-1)*D}=\ln(\boldsymbol{x}_{1*D})\boldsymbol{\Psi}^{\mathrm{T}}_{D*(D-1)}$，满足 $ILR(\boldsymbol{x})_{1*(D-1)}=\ln(\boldsymbol{x}_{1*D})\boldsymbol{\Psi}^{\mathrm{T}}_{D*(D-1)}$，其中 [1]，$\psi_{ij}=\begin{cases}+\sqrt{\frac{1}{(D-i)(D-i+1)}}, & j\le D-i\\ -\sqrt{\frac{D-i}{D-i+1}}, & j=D-i+1\\ 0, & \text{其他情况}\end{cases}$。其结构本质是通过一系列正交的组内或组间对数比构建而成。

性质 3：1) 正交性。矩阵行向量两两正交，即 $\boldsymbol{\Psi}\boldsymbol{\Psi}^{\mathrm{T}}=\boldsymbol{E}_{D-1}$。

2) 投影等价性。$\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi}=\boldsymbol{B}$，即 $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$ 与 $\boldsymbol{B}$ 是同一投影矩阵，ILR 变换可看作先通过 $\boldsymbol{B}$ 中心化投影，再通过 $\boldsymbol{\Psi}$ 提取正交基，实现无冗余降维。

事实上，三种对数比变换之间是可以相互转换的。

定理 3：$CLR(\boldsymbol{x})=ILR(\boldsymbol{x})\boldsymbol{\Psi}$，$ILR(\boldsymbol{x})=CLR(\boldsymbol{x})\boldsymbol{\Psi}^{\mathrm{T}}$；$ALR(\boldsymbol{x})=CLR(\boldsymbol{x})\boldsymbol{F}$，$CLR(\boldsymbol{x})=ALR(\boldsymbol{x})\boldsymbol{F}^{-}$；$ALR(\boldsymbol{x})=ILR(\boldsymbol{x})\boldsymbol{\Psi}\boldsymbol{F}$，$ILR(\boldsymbol{x})=ALR(\boldsymbol{x})\boldsymbol{F}^{-}\boldsymbol{\Psi}^{\mathrm{T}}$。

其中 $\boldsymbol{F}^{-}=\frac{1}{D}\begin{pmatrix}D-1 & -1 & -1 & \cdots & -1 & -1\\ -1 & D-1 & -1 & \cdots & -1 & -1\\ -1 & -1 & D-1 & \cdots & -1 & -1\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ -1 & -1 & -1 & \cdots & D-1 & -1\end{pmatrix}$，为摩尔－彭罗斯广义逆。

证明：给定 $\{\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_{D-1}\}$ 为单纯形 $\boldsymbol{S}^{D-1}$ 的一组标准正交基，则有

$$ILR(\boldsymbol{x})=\left(\langle\boldsymbol{x},\boldsymbol{e}_1\rangle_\alpha,\langle\boldsymbol{x},\boldsymbol{e}_2\rangle_\alpha,\cdots,\langle\boldsymbol{x},\boldsymbol{e}_{D-1}\rangle_\alpha\right)=(q_1,q_2,\cdots,q_{D-1}),\quad \boldsymbol{x}=\oplus_{i=1}^{D-1}q_i\odot\boldsymbol{e}_i,\quad q_i=\langle\boldsymbol{x},\boldsymbol{e}_i\rangle_\alpha。$$

又由定理 4，且 $\boldsymbol{\Psi}=\left(CLR(\boldsymbol{e}_1),CLR(\boldsymbol{e}_2),\cdots,CLR(\boldsymbol{e}_{D-1})\right)^{\mathrm{T}}$，可知

$$CLR(\boldsymbol{x})=CLR\left(\oplus_{i=1}^{D-1}q_i\odot\boldsymbol{e}_i\right)=q_1\cdot CLR(\boldsymbol{e}_1)+q_2\cdot CLR(\boldsymbol{e}_2)+\cdots+q_{D-1}\cdot CLR(\boldsymbol{e}_{D-1})=ILR(\boldsymbol{x})\boldsymbol{\Psi}$$

由 CLR 变换的保内积性，可得

$$\begin{aligned}ILR(\boldsymbol{x})&=\left(\langle\boldsymbol{x},\boldsymbol{e}_1\rangle_\alpha,\langle\boldsymbol{x},\boldsymbol{e}_2\rangle_\alpha,\ldots,\langle\boldsymbol{x},\boldsymbol{e}_{D-1}\rangle_\alpha\right)\\&=\left(\langle CLR(\boldsymbol{x}),CLR(\boldsymbol{e}_1)\rangle,\cdots,\langle CLR(\boldsymbol{x}),CLR(\boldsymbol{e}_{D-1})\rangle\right)\\&=CLR(\boldsymbol{x})\cdot\boldsymbol{\Psi}^{\mathrm{T}}\end{aligned}$$

[1]单纯形上不同正交基的相应变换矩阵并不相同，此处只给出了一种较为常见的情形。

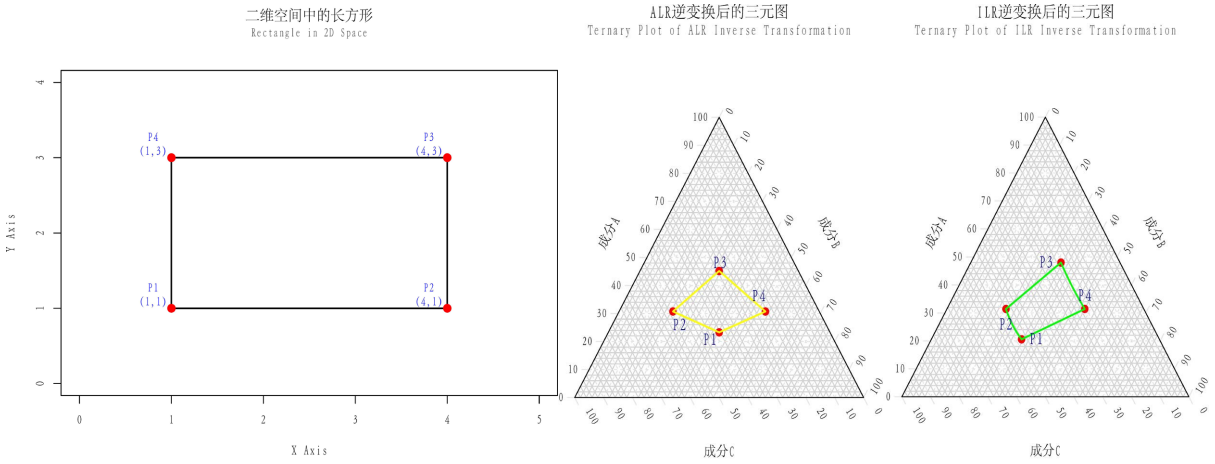
通过对数比的拆分，可将 ALR 变换的每个分量表示为 CLR 变换分量的线性组合，即 $ALR(x_i)=CLR(x_i)-CLR(x_D), i=1,2,\cdots,D-1$，即 $ALR(\boldsymbol{x})=CLR(\boldsymbol{x})\boldsymbol{F}$。

进一步可得：$CLR(\boldsymbol{x})=ALR(\boldsymbol{x})\boldsymbol{F}^{-}$。

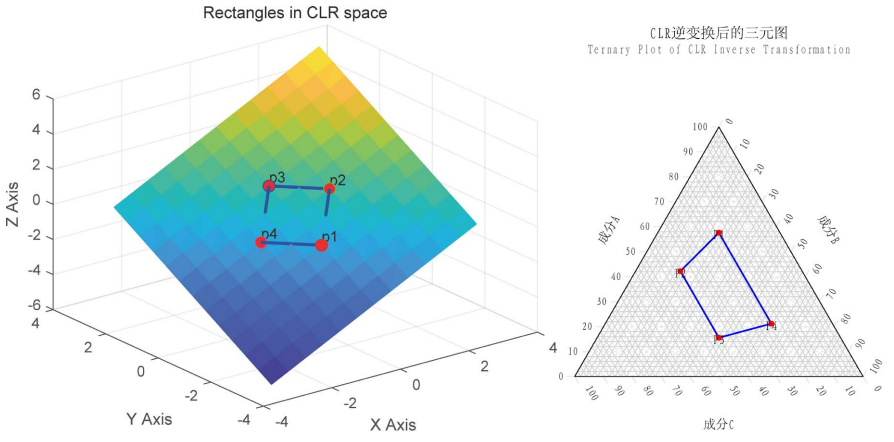
3. 三种对数比变换的一致性

(一) 三种对数比变换的三元图示几何意义一致性

以下给出欧氏空间中常见长方形及椭圆通过三种对数比变换后的三元图图示结果。



(a) 二维长方形　(b) ALR 逆变换　(c) ILR 逆变换



(d) 三维长方体　(e) CLR 逆变换

Figure 1. Three kinds of log-ratio inverse transformations of rectangles in Euclidean space
图 1. 欧氏空间中矩形的三种对数比逆变换

如图 1、图 2 所示，三元图中无法呈现出三种对数比的保角性和保距性等性质。为解决这一问题，需要借助从单纯形空间到欧式空间的三种变换，进而利用变换后的欧氏空间图形空间关系进行相应空间位置关系确定。

基于欧氏空间中两个向量夹角与距离的几何意义可知三元图中两个成分向量夹角反映相应成分向量变化方向的相似性，距离衡量相应成分向量的相似度。

以一个具体事实说明两成分向量夹角和距离的意义。假设研究某饮料中 3 种基础成分(果汁 A、糖浆 B、水 C)的三款配方，成分占比之和均为 1。如表 1、表 2 所示。

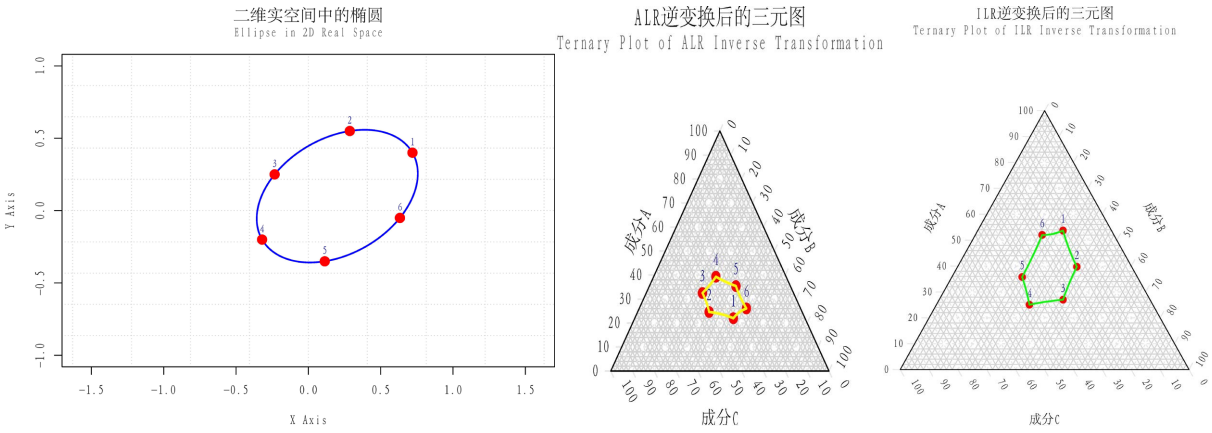
Table 1. Original beverage formula
表 1. 饮料原配方

基础配方	A	B	C
配方 1 (常规款)	0.3	0.2	0.5
配方 2 (微调款)	0.32	0.18	0.5
配方 3 (差异款)	0.1	0.4	0.5

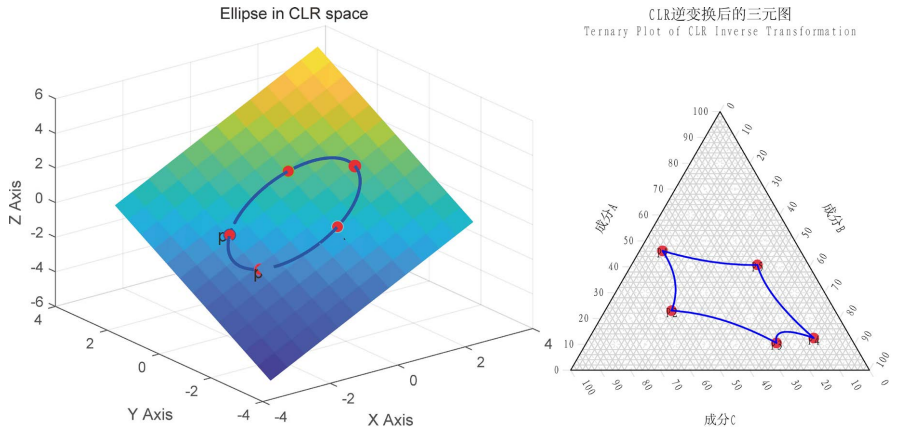
Table 2. Modified formula based on “Formula 1”
表 2. 基于“配方 1”的改良配方

改良配方	A	B	C
配方 X (增甜)	0.25	0.25	0.5
配方 Y (增浓)	0.35	0.15	0.5
配方 Z (微调增甜)	0.28	0.15	0.5

三种对数比变换两向量夹角大小皆反映了相应成分向量比例变化的相似性。以 *ILR* 变换为例，*ILR* 变换后$\langle X,Y\rangle \approx 179^{\circ}$，由表 2 与表 1 对比可以看出，配方 X 进行减 A 加 B，配方 Y 。



(a) 二维椭圆　(b) ALR 逆变换　(c) ILR 逆变换



(d) 三维球体　(e) CLR 逆变换

Figure 2. Three kinds of log-ratio inverse transformations of ellipses in Euclidean space
图 2. 欧式空间中椭圆的三种对数比逆变换

进行加 A 减 B，变化方向完全相反，因此变换后两向量角度越大，变化方向越差异；同理，$\langle X,Z\rangle \approx 11^{\circ}$，两向量夹角越小，变化方向越一致。三种对数比变换后两向量距离大小与相应成分向量比例结构相似度有关。以 *ALR* 变换为例，*ALR* 变换后 $d_{\alpha}\langle 1,2\rangle \approx 0.41$，由表 1 可以看出，配方 2 是对配方 1 的微调，因此两者距离越小，比例结构越相似；同理，$d_{\alpha}\langle 1,3\rangle \approx 1.47$，两成分向量距离越大，差异越显著。

(二) 三种对数比变换的线性组合

设 *LR* 代表三种对数比变换 *ALR*，*CLR*，*ILR* 中的任意一种。

定理 4：对于成分项 $\boldsymbol{x}=(x_1,\cdots,x_D)^{\mathrm{T}}, \boldsymbol{y}=(y_1,\cdots,y_D)^{\mathrm{T}} \in S^{D-1}$ 以及实数 α,β，有

$$LR(\alpha\odot x\oplus\beta\odot y)=\alpha LR(x)+\beta LR(y)$$

证明：

$$ALR(\alpha\odot x\oplus\beta\odot y)=ALR\left(c\left(x_1^{\alpha}y_1^{\beta},x_2^{\alpha}y_2^{\beta},\cdots,x_D^{\alpha}y_D^{\beta}\right)\right)$$

$$=\left(\alpha\ln\frac{x_1}{x_D}+\beta\ln\frac{y_1}{y_D},\alpha\ln\frac{x_2}{x_D}+\beta\ln\frac{y_2}{y_D},\cdots,\alpha\ln\frac{x_{D-1}}{x_D}+\beta\ln\frac{y_{D-1}}{y_D}\right)$$

$$=\alpha ALR(x)+\beta ALR(y)$$

$$ILR(\alpha\odot x\oplus\beta\odot y)=\alpha\left((x,e_1)_{\alpha},(x,e_2)_{\alpha},\cdots,(x,e_{D-1})_{\alpha}\right)+\beta\left((y,e_1)_{\alpha},(y,e_2)_{\alpha},\cdots,(y,e_{D-1})_{\alpha}\right)$$

$$=\alpha ILR(x)+\beta ILR(y)$$

CLR 证明与 *ALR* 同理，这里不再展现。

4. 三种对数比变换的应用

ALR 适用于具有明确基准成分的统计分析，突出了其他成分相对基准成分的变化，但其缺乏对称性，不具等距性、保角性，且基准成分的选择对分析结果至关重要。以古代玻璃文物化学成分数据为例，对相应成分向量进行 *ALR*、*CLR*、*ILR* 3 种不同的对数比变换，基于 KMO 检验和 Bartlett 球形度检验只有 *ALR* 后的数据更适合做主成分分析，且获得的结果更加准确，此研究过程为成分向量的主成分分析提供了有效的借鉴思路和方法[3]。

CLR 适用于描述成分向量的整体构成，整体变化情况，满足对称性，但变换后向量数据受制零和限制，存在共线性，不适宜进行多元回归分析，主要用于探索性数据分析。为提取湘西北铅锌矿所在区域的水系沉积物常量元素组合异常，采用 *CLR* 等多种对数比变换对沉积物原始数据预处理，再进行偏最小二乘降维分析。尽管该研究中指出等距对数比变换效果更优，但 *CLR* 变换同样有效解决了元素含量数据的闭合效应问题，为后续异常提取排除了伪相关干扰，且变换后的数据能更好适配多元统计模型[4]。

由于 *ILR* 能保持成分向量的空间形态，不受定和限制，适用于多元成分向量的降维及回归、聚类、因子分析等场景，但不具对称性且结果解释较为复杂。在区域化探数据分析中，*ILR* 可以有效构建化探数据的标准正交基，消除其闭合效应，解释数据的组成性质，为区域化物探数据的定量分析和找矿预测提供了更为可靠的理论支撑[5]。

5. 总结

成分数据在地质学、生态学、食品科学等众多领域中应用广泛，但受限于“定和”约束，传统多元统计分析方法难以直接应用。本文系统梳理了三种对数比变换的差异性与一致性。在差异性方面，从几何特性(等距性、保角性)与数学表达(矩阵表示及转换)两个维度，揭示了不同变换的区别；在一致性方面，以三元成分向量为例(可扩展到三元以上)，通过三元图呈现不同变换对原始数据相对结构的特性保留，并验证了三类变换在数据线性关系传递上的共性规律。同时，本文进一步结合具体研究场景，明确了不同

对数比变换的适用条件与选择依据。本文对成分数据的理论研究与应用实践皆具有一定的参考价值与指导意义。

基金项目

大学生创新训练项目“成分数据初探”(项目编号：202507011)。

参考文献

[1] Aitchison, J. (1986) The Statistical Analysis of Compositional Data. Chapman and Hall, London.

[2] Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G. and Barceló-Vidal, C. (2003) Isometric Logratio Transformations for Compositional Data Analysis. *Mathematical Geology*, **35**, 279-300. https://doi.org/10.1023/a:1023818214614

[3] 王雪, 谢淼, 周玲菲, 等. 基于成分向量处理的主成分分析研究[J]. 科学技术创新, 2023(18): 94-98.

[4] 王琨, 肖克炎, 丛源. 对数比变换和偏最小二乘法在地球化学组合异常提取中的应用——以湘西北铅锌矿为例[J]. 物探与化探, 2015, 39(1): 141-148.

[5] 李柱, 张德会, 杨帆, 等. 等距对数比变换及混合分布在区域化探数据分析中的应用[J]. 现代地质, 2023, 37(3): 662-673.