Performing Factor Analysis with ilr and clr coordinates

Homework 4 DS4GS

Problem statement

- **ilr coordinates** form an orthonormal isometric basis, allows applying multi-variate statistical methods
- But: Impossible to interpret loadings to ilr coordinates in terms of original elements (essentially non-uniqueness of the basis)
- **clr coordinates** are isometric, but not orthogonal
- But: loading to clr coordinates can be interpreted in terms of elements

Solution

- Factor analysis requires a covariance matrix
- Covariance of ilr is not singular, but cannot be interpreted in terms of elements
- Covariance of clr is singular, but can be interpreted in terms of elements
- Build a projection between ilr and clr covariance matrices that is not singular and can be interpreted in terms of original covariance

Math & workflow

step1: scale ilr basis *e onto* clr basis

Build the ilr basis in the matrix M of size $p \times p-1$, that is

$$M = [\vec{e}_1 \quad \vec{e}_2 \quad \cdots \quad \vec{e}_{p-1}]$$

with $\vec{e}_i = \Bigg(\underbrace{\sqrt{\frac{1}{i(i+1)}}, \ldots, \sqrt{\frac{1}{i(i+1)}}}_{i \text{ elements}}, -\sqrt{\tfrac{i}{i+1}}, 0, \ldots, 0\Bigg)$ the i-th column.

Compute the standard deviation for the columns of X_c. This is a vector $s = (s_1, s_2, \ldots, s_p)$ of size p. Construct the projection

$$P = \begin{bmatrix} \frac{1}{s_1} & & & \\ & \frac{1}{s_2} & & \\ & & \ddots & \\ & & & \frac{1}{s_p} \end{bmatrix} M$$

step2: calculate the covariance of the projection

$$C = XX^T \Rightarrow$$

$$C_{projection} = PX(PX)^T = PXX^TP^T = PCP^T$$

Overview

Given the data matrix X of size $n \times p$ with n the number of samples and p the number of components:

- Perform clr and ilr transformations to produce X_i and X_c.
- Build the ilr basis in the matrix M of size $p \times p-1$, that is

$$M = \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \cdots & \vec{e}_{p-1} \end{bmatrix}$$

with $\vec{e}_i = \Bigg(\underbrace{\sqrt{\frac{1}{i(i+1)}}, \ldots, \sqrt{\frac{1}{i(i+1)}}}_{i \text{ elements}}, -\sqrt{\tfrac{i}{i+1}}, 0, \ldots, 0 \Bigg)$ the i-th column.

- Compute the standard deviation for the columns of X_c. This is a vector $s = (s_1, s_2, \ldots, s_p)$ of size p. Construct the projection

$$P = \begin{bmatrix} \frac{1}{s_1} & & & \\ & \frac{1}{s_2} & & \\ & & \ddots & \\ & & & \frac{1}{s_p} \end{bmatrix} M$$

- Compute the covariance in ilr coordinates C_i of size $p-1 \times p-1$ from the matrix X_i.
- Compute the final covariance $C = P C_i P^\top$ of size $p \times p$ for factor analysis.