# Geometry of the simplex: modelling of processes

**J. J. Egozcue**

Dep. Matemática Aplicada III
UPC, Barcelona
juan.jose.egozcue@upc.edu

Facultad de Ciencias; Departamento de Matemáticas
Universidad de los Andes, Bogotá
15 de Marzo, 2007

# Summary

# Mathematical sentences

- All Euclidean spaces of equal dimension are isomorphic and isometric
- All separable Hilbert spaces of equal dimension are isomorphic and isometric

**Consequence**

If $\mathcal{S}$ is a set and there is a one-to-one mapping $\varphi : \mathcal{S} \to \mathbb{R}^n$ then an Euclidean structure is induced on $\mathcal{S}$

**Which particular $\varphi$ is adequate?**

Operations and metrics should be interpretable and applicable

# Mathematical sentences

- All Euclidean spaces of equal dimension are isomorphic and isometric
- All separable Hilbert spaces of equal dimension are isomorphic and isometric

**Consequence**

If $\mathcal{S}$ is a set and there is a one-to-one mapping $\varphi : \mathcal{S} \to \mathbb{R}^n$ then an Euclidean structure is induced on $\mathcal{S}$

**Which particular $\varphi$ is adequate?**

Operations and metrics should be interpretable and applicable

# The case of the simplex of $n$ parts

$$\mathcal{S}^n = \left\{ \mathbf{x} = [x_1, \ldots, x_n] \in \mathbb{R}^n \;\middle|\; x_i > 0, \sum_{i=1}^{n} x_i = \kappa \right\}$$

The way was not to find $\varphi$ !

- non-isometric transformations: alr, clr (Aitchison 1982-86)
- perturbation (sum) (Aitchison, 1982)
- distance, clr-compatible (Aitchison, 1982)
- Euclidean structure (Billheimer et al. 2001; Pawlowsky-Glahn and Egozcue 2001)
- isometry, $\varphi$ (Egozcue et al. 2003)

# Euclidean space structure of $\mathcal{S}^n$

for $\mathbf{x}, \mathbf{y} \in \mathcal{S}^n$, $\alpha \in \mathbb{R}$, and $\mathcal{C}$ the closure operation

- perturbation: $\mathbf{x} \oplus \mathbf{y} = \mathcal{C}[x_1 y_1, \ldots, x_n y_{nn}]$
- powering: $\alpha \odot \mathbf{x} = \mathcal{C}[x_1^\alpha, \ldots, x_n^\alpha]$
- inner product:

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{n} \sum_{i<j} \ln \frac{x_i}{x_j} \ln \frac{y_i}{y_j}$$

- associated norm and distance:

$$\|\mathbf{x}\|_a^2 = \frac{1}{n} \sum_{i<j} \left( \ln \frac{x_i}{x_j} \right)^2 \quad d_a^2(\mathbf{x}, \mathbf{y}) = \frac{1}{n} \sum_{i<j} \left( \ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j} \right)^2$$

# Euclidean space structure of $\mathcal{S}^n$

for $\mathbf{x}, \mathbf{y} \in \mathcal{S}^n$, $\alpha \in \mathbb{R}$, and $\mathcal{C}$ the closure operation

- perturbation: $\mathbf{x} \oplus \mathbf{y} = \mathcal{C}[x_1 y_1, \ldots, x_n y_{nn}]$
- powering: $\alpha \odot \mathbf{x} = \mathcal{C}[x_1^\alpha, \ldots, x_n^\alpha]$
- inner product:

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{n} \sum_{i<j} \ln \frac{x_i}{x_j} \ln \frac{y_i}{y_j}$$

- associated norm and distance:

$$\|\mathbf{x}\|_a^2 = \frac{1}{n} \sum_{i<j} \left( \ln \frac{x_i}{x_j} \right)^2 \quad d_a^2(\mathbf{x}, \mathbf{y}) = \frac{1}{n} \sum_{i<j} \left( \ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j} \right)^2$$

# Example of orthogonal coordinates (using SBP)

A difference with $\mathbb{R}^n$: a canonical basis is not defined
Alternative: orthonormal basis of balances

| level | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $r$ | $s$ |
|---|---|---|---|---|---|---|---|---|
| 1 | $+1$ | $+1$ | $-1$ | $-1$ | $+1$ | $+1$ | 4 | 2 |
| 2 | $+1$ | $-1$ | 0 | 0 | $-1$ | $-1$ | 1 | 3 |
| 3 | 0 | $+1$ | 0 | 0 | $-1$ | $-1$ | 1 | 2 |
| 4 | 0 | 0 | 0 | 0 | $+1$ | $-1$ | 1 | 1 |
| 5 | 0 | 0 | $-1$ | $+1$ | 0 | 0 | 1 | 1 |
| 1 | $+\frac{1}{\sqrt{12}}$ | $+\frac{1}{\sqrt{12}}$ | $-\frac{1}{\sqrt{3}}$ | $-\frac{1}{\sqrt{3}}$ | $+\frac{1}{\sqrt{12}}$ | $+\frac{1}{\sqrt{12}}$ | | |
| 2 | $+\frac{\sqrt{3}}{2}$ | $-\frac{1}{\sqrt{12}}$ | 0 | 0 | $-\frac{1}{\sqrt{12}}$ | $-\frac{1}{\sqrt{12}}$ | | $\Psi$ |
| 3 | 0 | $+\frac{\sqrt{2}}{\sqrt{3}}$ | 0 | 0 | $-\frac{1}{\sqrt{6}}$ | $-\frac{1}{\sqrt{6}}$ | | |
| 4 | 0 | 0 | 0 | 0 | $+\frac{1}{\sqrt{2}}$ | $-\frac{1}{\sqrt{2}}$ | | |
| 5 | 0 | 0 | $+\frac{1}{\sqrt{2}}$ | 0 | 0 | $-\frac{1}{\sqrt{2}}$ | | |

# The ilr isometry

Properties of $\Psi$

- $\Psi$ is a $(n-1, n)$-matrix
- $\Psi\Psi^t = I_{n-1}$
- $\Psi^t\Psi = I_n - (1/n)\mathbf{1}_n^t\mathbf{1}_n$

From coordinates to compositions

$$\mathbf{x} = \bigoplus_{i=1}^{n-1} x_i^* \odot \mathbf{e}_i = \mathcal{C}\exp(\mathbf{x}^*\Psi)\ ,\ x_i^* = \langle \mathbf{x}, \mathbf{e}_i \rangle_a\ ,$$

From compositions to coordinates

$$\mathbf{x}^* = \ln(\mathbf{x}) \cdot \Psi^t\ ,$$

# Linear transformation (endomorphism)

Linear transformation in $\mathbb{R}^{n-1}$ (coordinates): for any $A^*$,

$$\mathbf{y}^* = \mathbf{x}^* A^*$$

Linear transformation in $\mathcal{S}^n$, applying $\varphi^{-1}$

$$\mathbf{y} = \mathbf{x} \circ A = \mathcal{C}\Big[\ldots, \underbrace{\prod_{k=1}^{n} x_k^{a_{k\ell}}}_{\ell-\text{term}}, \ldots\Big]$$

$A$ is not a general $(n, n)$-matrix:

- $A^* = \Psi A \Psi^t \quad , \quad A = \Psi^t A^* \Psi$
- Rows and columns add to 0
- Rank($A$)=Rank($A^*$) $\leq n-1$

# Exponential growth or decay follow straight-lines

The mass of $n$ radioactive isotopes decay with rates $\lambda_i$ ($\lambda_i \leq 0$ in this case) as

$$x_i(t) = x_i(0) \cdot \exp(\lambda_i t)$$

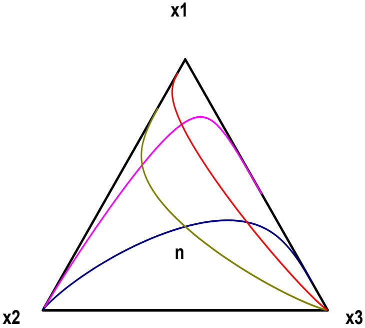
Composition of the sample at time $t$ is

$$\mathcal{C}\mathbf{x}(t) = \underbrace{\mathbf{x}(0)}_{\text{origin}} \oplus (t \odot \underbrace{\exp[\boldsymbol{\lambda}]}_{\text{direction}})$$

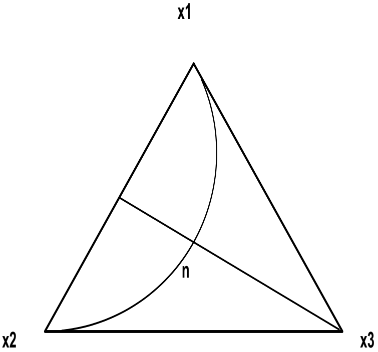
# Compositional lines

Correspond to exponential growth or decay of masses

$$\mathbf{y} = \mathbf{x}_0 \oplus (\alpha \odot \mathbf{x}_1)$$

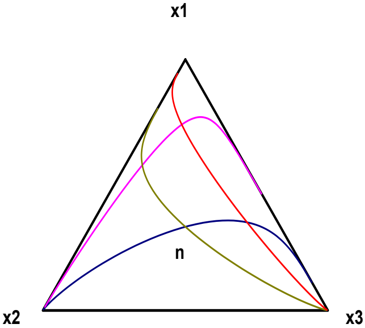



parallel lines

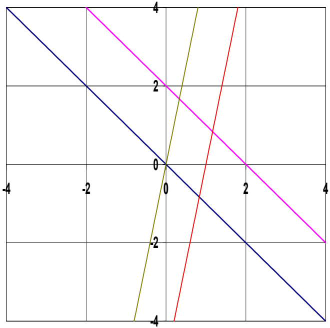



orthogonal lines

# Compositional lines in coordinates



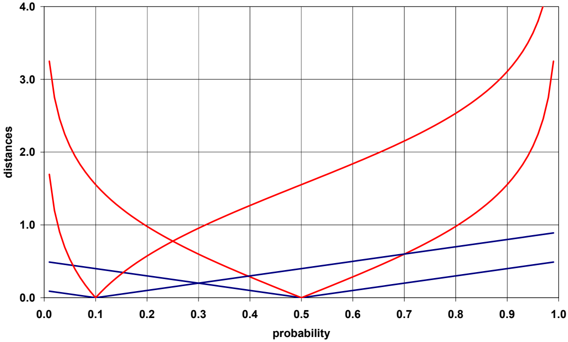

in $\mathcal{S}^3$




coordinate representation

# Distance in $\mathcal{S}^2$

Distance of a point in $\mathcal{S}^2$ to a reference ($\mathbb{R}$, $\mathcal{S}^2$).
End-points are at the infinity for $\mathcal{S}^2$



# Aitchison measure in $\mathcal{S}^n$

Lebesgue measure of a hyper-rectangle in $\mathbb{R}$ and $\mathbb{R}^{n-1}$ (Cartesian coordinates)

$$\lambda\{(a_i, b_i)\} = |b_i - a_i| \ , \ \lambda_{n-1}\{(\mathbf{a}, \mathbf{b})\} = \prod_{i=1}^{n-1} |b_i - a_i|$$

For a general borelian, $B^*$, $\lambda\{B^*\}$.

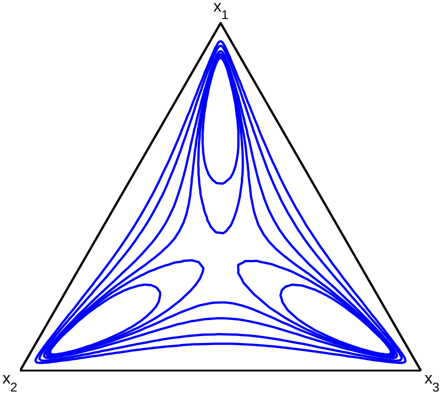
Aitchison measure of $B = \varphi^{-1}(B^*)$ in $\mathcal{S}^n$
Coordinates (isometry): $\mathbf{x}^* = \varphi(\mathbf{x})$

$$\mathbf{x} \in \mathcal{S}^n \ , \ \mathbf{x}^* \in \mathbb{R}^{n-1}$$

$$\lambda_{\mathcal{S}^n}\{B\} = \lambda_{n-1}\{\varphi(B)\} = \lambda_{n-1}\{B^*\}$$

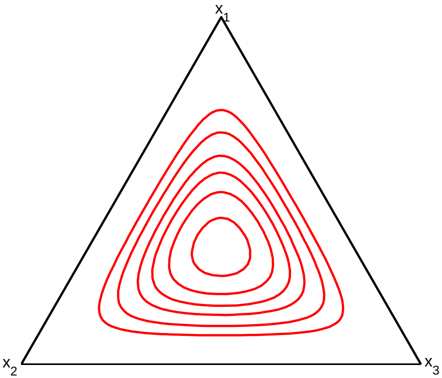
# Normal on the simplex (logistic-normal)

$\mathcal{S}^3 \subset \mathbb{R}^2$, Lebesgue measure as reference:
Radon-Nikodym derivative: $f = \frac{dP}{d\lambda}$



# Normal on the simplex (logistic-normal)

$\mathcal{S}^3$ as Euclidean space, Aitchison measure as reference:
Radon-Nikodym derivative: $f = \frac{dP}{d\lambda_{\mathcal{S}}} = \frac{dP}{d\lambda} \cdot \frac{d\lambda}{d\lambda_{\mathcal{S}}}$



# Consequences and extension

Aitchison Euclidean geometry in the simplex motivates:

- Re-definition of measures of location and dispersion in statistics
- An extension from compositions to probability densities (and measures)
- The calculus in the simplex: new differential models

# Derivative

If differences in $\mathcal{S}^n$ are computed using $\ominus \equiv (-1)\odot$, definition of derivatives should change.

$$\mathbf{h} : \mathbb{R} \longrightarrow \mathcal{S}^n$$

Derivative in the simplex

$$\begin{aligned} D^{\oplus}\mathbf{h}(t) &= \lim_{\tau \to 0} \frac{1}{\tau} \odot (\mathbf{h}(t+\tau) \ominus \mathbf{h}(t)) \\ &= \varphi^{-1} \frac{d}{dt} \varphi(\mathbf{h}(t)) \\ &= \mathcal{C} \exp \frac{d}{dt} \ln(\mathbf{h}(t)) \end{aligned}$$

# Example: predator-prey Volterra ODE

$$D_t x_1 = b_1 x_1 + a_{12} x_1 x_2 \qquad \frac{D_t x_1}{x_1} = b_1 + a_{12} x_2$$

$$D_t x_2 = b_2 x_2 + a_{21} x_1 x_2 \qquad \frac{D_t x_2}{x_2} = b_2 + a_{21} x_1$$

Since $D^{\oplus}(\cdot) = \mathcal{C} \exp D_t \ln(\cdot)$,

$$D^{\oplus}\mathbf{x} = \exp[b_1, b_2] \oplus \exp(\mathbf{x}) \circ \begin{pmatrix} 0 & a_{21} \\ a_{12} & 0 \end{pmatrix}$$

or in coordinates (a single ODE)

$$D_t \varphi(\mathbf{x}) = \varphi(\exp[b_1, b_2]) + \varphi(\exp(\mathbf{x}))A^*$$

$$A^* = -\frac{a_{12} + a_{21}}{2}$$

# Example: a linear system in $\mathcal{S}^3$

Expression in $\mathcal{S}^3$

$$D^{\oplus}\mathbf{x} = \mathbf{x} \circ \begin{pmatrix} 0.015 & -0.092 & 0.082 \\ 0.861 & -0.358 & -0.216 \\ -0.871 & 0.419 & 0.161 \end{pmatrix}$$

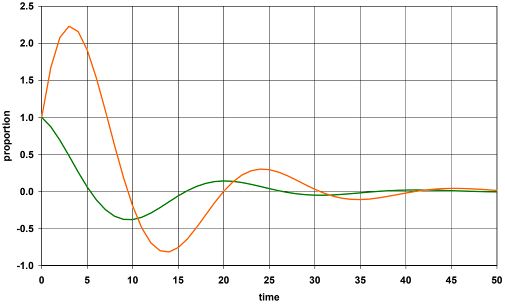
$$\mathbf{x}(0) = \mathcal{C}\left[\sqrt{\frac{2}{3}}, \frac{\sqrt{3}-1}{\sqrt{6}}, -\frac{\sqrt{3}+1}{\sqrt{6}}\right]$$

Expression in coordinates ($\mathbb{R}^2$)

$$D_t\varphi(\mathbf{x}) = \varphi(\mathbf{x}) \begin{pmatrix} 0.01 & -0.1 \\ 1 & -0.2 \end{pmatrix} \quad , \quad \mathbf{x}(0) = [1, 1]$$

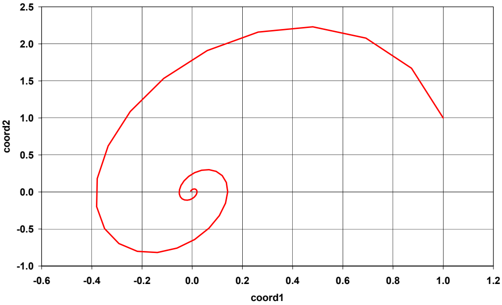
# Solution of the system

coordinates in time



# Solution of the system

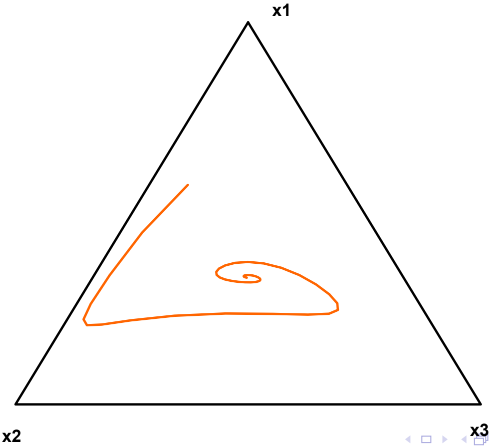
phase in coordinates



Calculus

# Solution of the system

phase in $\mathcal{S}^3$




V. Pawlowsky-Glahn
and
J. J. Egozcue

# Simplicial integrals

Sum is not allowed in the simplex. Perturbation used instead.

$$\mathbf{h} : \mathbb{R} \longrightarrow \mathcal{S}^n$$

$$\int^{\oplus} \mathbf{h}(t)\, dt = \varphi^{-1}\left(\int \varphi(\mathbf{h}(t))\, dt\right) = \mathcal{C} \exp\left(\int \ln(\mathbf{h}(t))\, dt\right)$$

Riemann sums are

$$\int^{\oplus} \mathbf{h}(t)\, dt \approx \bigoplus_i (t_{i+1} - t_i) \odot \mathbf{h}(t_i')$$

# Example: integration of a $\mathcal{S}^n$ valued function

Linear function: (value of a portfolio; disintegration; growth of a population)

$$\mathbf{h}(t) = \mathbf{x}_0 \oplus (t \odot \exp(\boldsymbol{\lambda}))$$

Problem: find an average value of $\mathbf{h}(t)$ in $(0, T)$
Naive approach:

$$\tilde{\mathbf{h}} = \frac{1}{T} \cdot \int_{(0,T)} \mathbf{h}(t)\, dt$$

Simplicial mean value:

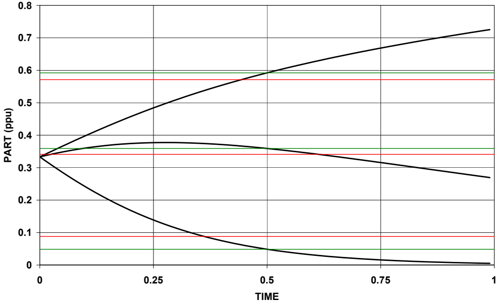
$$\bar{\mathbf{h}} = \frac{1}{T} \odot \int_{(0,T)}^{\oplus} \mathbf{h}(t)\, dt = \phi^{-1}\left(\frac{1}{T} \cdot \int_{(0,T)} \mathbf{h}^*(t)\, dt\right)$$

# Example of mean value in $(0, T = 1)$

$n = 3$, $\boldsymbol{\lambda} = [0, -1, -5]$, $\mathbf{x}_0 = \mathcal{C}[1, 1, 1]$

Naive approach: mean is not a point of the process(!!)

Simplicial app.: corresponds to value at $T/2$



# Conclusions

- Aitchison geometry of the simplex provides a specific calculus
- Simplicial differential and integral operators may be useful in modelling compositional phenomena

# Further reading and activities

- Mathematical Geology Vol. 37 Nr. 7 (2005) – special issue on compositional data analysis
- Compositional data analysis in the Geosciences: From theory to practice (October 2006) — special publication of the Geological Society (SPE 264)
- CoDaWork'08, Girona (Spain), May 2008 (http://ima.udg.es/Activitats/CoDaWork08/)

## References

Aitchison, J. (1982): The statistical analysis of compositional data (with discussion). *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **44**, 2, 139–177.

Aitchison, J. (1986): *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). (Reprinted in 2003 with additional material by The Blackburn Press). 416 p.

Pawlowsky-Glahn, V. and Egozcue, J. J.(2001): Geometric Approach to Statistical Analysis on the Simplex, *Stochastic Environmental Research and Risk Assessment*, **15**, 5, 384-398.

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G. and Barceló-Vidal C.(2003): Isometric logratio transformations for compositional data analysis, *Mathematical Geology*, **35**, 3, 279-300.

Egozcue, J. J. and Pawlowsky-Glahn, V. (2005): Groups of parts and their balances in compositional data analysis, *Mathematical Geology*, **37**, 7, 799-832.