Using log-constrast relations to generate Normal variables on the $\mathcal{S}$implex

Universitat de Girona
Departament d'Informàtica i Matemàtica Aplicada

M. Comas G. Mateu-Figueras S. Thió-Henestrosa
Universitat de Girona

Abstract

It is well known that the covariance structure of a D-part composition is completely determined by $\frac{D(D-1)}{2}$ *independent* covariances between log-contrasts[1]. In this poster, a sufficient condition is given showing when a covariance structure is completely determined by a given amount of relations. Moreover, it is shown how the characterization is applied into CoDaPack software. The routine iteratively gets information from the user via constant log-contrast and log-ratio relations present in the distribution, when the distribution is completely determined the routine generate a random sample with the specified conditions.

1.- Theoretical basis

Let $e = D(D-1)/2$. Let $\mathcal{B}$ be a base from CLR-space. Consider the application $\phi_{\mathcal{B}}$ from $\mathbb{R}^D \times \mathbb{R}^d$ to $\mathbb{R}^e$ defined by

$$\phi_{\mathcal{B}}(a, a') = \begin{pmatrix} \varphi_1 {\varphi'}_1 + \varphi_1 {\varphi'}_1 \\ \varphi_1 {\varphi'}_2 + \varphi_2 {\varphi'}_1 \\ \vdots \\ \varphi_i {\varphi'}_j + \varphi_j \varphi'_i \\ \vdots \\ \varphi_D {\varphi'}_D + \varphi_D {\varphi'}_D \end{pmatrix} \text{ for } i \leq j$$

where φ and φ' are the coordinates of a and a' with respect to $\mathcal{B}$.

Then,

Defining the covariance structure

A set of m log-contrasts

$$\{a^1 \log(\mathbf{x}), a^2 \log(\mathbf{x}), \ldots a^m \log \mathbf{x}\}$$

and n relations of the form

$$\mathrm{Cov}(a^{l_i} \log(\mathbf{x}), a^{r_i} \log(\mathbf{x})) = c_i, \quad 1 \leq i \leq n$$

define the covariance structure if the set

$$\{\phi_{\mathcal{B}}(a^{l_i}, a^{r_i}) \in \mathbb{R}^e\}_{1 \leq i \leq n}$$

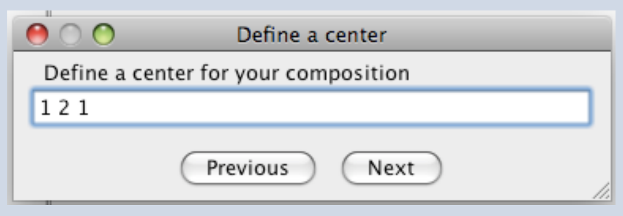
is a generator system of $\mathbb{R}^e$.

The following two conditions are proposed to characterize a normal distribution.

- Constant log-constrast relations.
  - The constant log-contrast is assumed independent to any orthogonal log-contrast.
- Bounded log-ratio relations.
  - Given a bound relation between components, a variance relation is implicitly given assuming certain level of confidence.

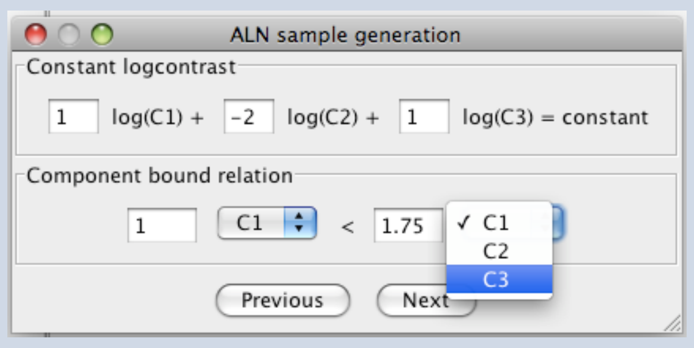
2.- CoDaPack routine: getting information from the user

Set the parameters of a normal distribution on $\mathcal{S}^D$ is straightforward for people used to compositional data transformations (i.e. set the parameter of a normal distribution on the coordinate space and transforming back to the Simplex space).
In the proposed routine for CoDaPack [2], the center of the composition is asked to the user via a simple menu.



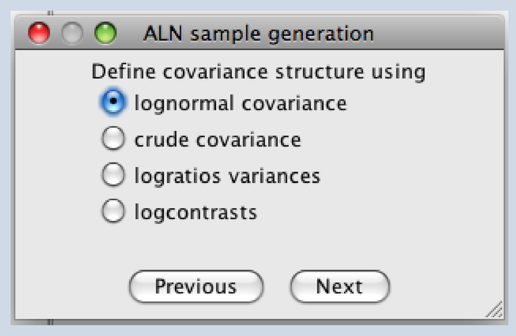
Asking for the center of a composition

The covariance structure is defined via *constant log-contrasts* and *bounds* between its components.



Getting log-contrast information from users

Other methods to define the covariance structure



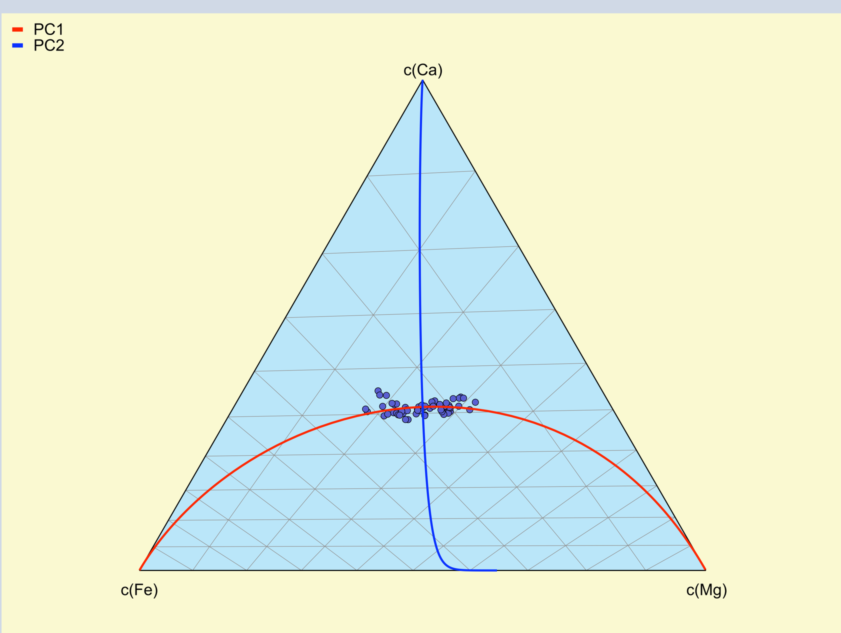
CoDaPack allow different methodologies to define the covariance structure: lognormal covariance, crude covariance, log-ratio covariance or present explained log-contrast relation.

CoDaPack is available at http://ima.udg.edu/codapack

3.- Example: Gabriele's Ankerites

Gabriele is a geologist interested in obtaining a random sample of Ankerites. He has a well formed knowledge about some properties of his data formed by Calcium Ca, Magnesium Mg and Iron Fe.
His sample, with center $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$, has the following equilibrium

$$\frac{Mg\ Fe}{Ca^2} \approx \text{constant.}$$



Gabriele's sample

In his sample, also the following relation holds

$$\frac{Fe}{Mg} \leq 1.75. \qquad (1)$$

With this information, Gabriele is interested in generate a random sample according to his expert knowledge. According to Gabriele's information the following relations holds:

$$\mathrm{Var}\left(\log(Mg) - 2\log(Ca) + \log(Fe)\right) \approx \epsilon$$

and

$$\mathrm{Var}\left(\log(Fe) - \log(Mg)\right) = \left(\frac{\log(1.75) - \mathbb{E}(\log(\frac{Fe}{Mg}))}{z_{\alpha/2}}\right)^2$$

with a confidence of $1 - \alpha$ that Eq.1 holds.
Finally, assuming the constant log-contrast

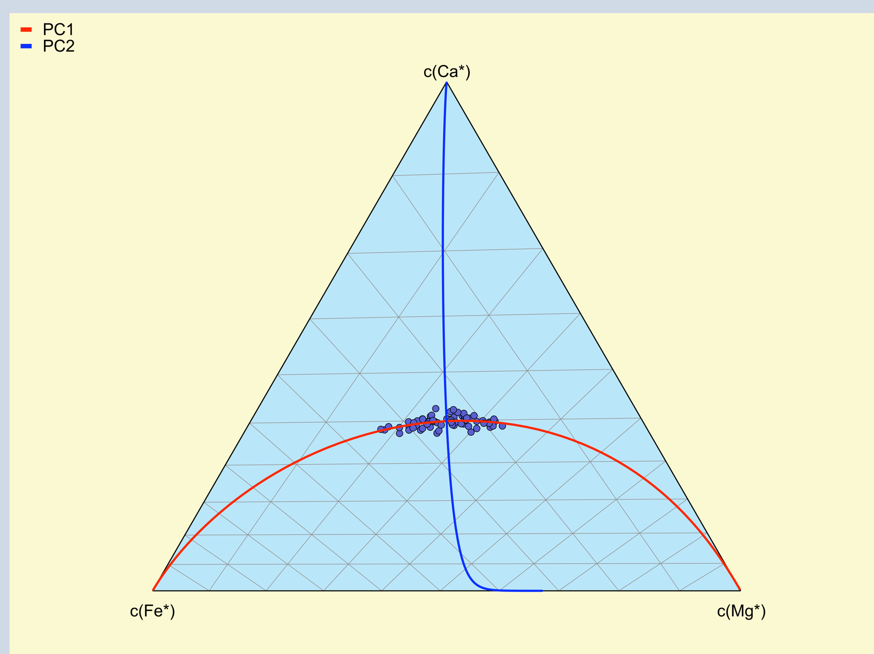
$$\log(Mg) - 2\log(Ca) + \log(Fe)$$

is independent to any orthogonal log-contrast $a_{\log}$, we also get

$$\mathrm{Cov}\left(\log(Mg) - 2\log(Ca) + \log(Fe), a_{\log}\right) = 0.$$

Given information

- $\mathbb{E}_{\mathcal{A}}\left[(Mg, Ca, Fe)\right] = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$
- $\mathrm{Var}\left(\log(Mg) - 2\log(Ca) + \log(Fe)\right) \approx \epsilon$
- $\mathrm{Var}\left(\log(Fe) - \log(Mg)\right) = \left(\frac{\log(1.75) - \mathbb{E}(\log(\frac{Fe}{Mg}))}{z_{\alpha/2}}\right)^2$
- $\mathrm{Cov}\left(\log(Mg) - 2\log(Ca) + \log(Fe), a_{\log}\right) = 0.$



Generated data using Gabriele's information

Conclusions and future work

The covariance structure of a distribution on $\mathcal{S}^D$ can be defined via logarithmic transformation (i.e. ILR, ALR, CLR [1, 3]) but the covariance matrix used to define the parameters could be difficult to interpret. In our approach, the covariance structure is not defined via an specific transformation but using log-contrast relations. In the Ankerites case, the covariance structure is completely determined given three different relations between log-contrast.

The log-contrast characterization defines when a set of relation completely determine the covariance structure. An obvious question is what must be done if not all the information is available. May the partial information be useful for specifying a prior distribution in bayesian inference?

References

[1] Aitchison, J, 2003, The statistical analysis of compositional data: (Reprint) Blackburn Press, Caldwell, NJ, 416p.

[2] Comas M, Thió-Henestrosa, S, CoDaPack 2.0: a stand-alone, Multi-platform Compositional Software. CoDaWork'11: 4th International Workshop on Compositional Data Analysis, Sant Feliu de Guíxols 2011

[3] Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G and Barceló-Vidal, C, 2003, Isometric logratio transformations for compositional data analysis. Math. Geol 39(11): 279-300

[4] Thió-Henestrosa, S, Martín-Fernández, JA, 2005, Dealing with compositional data: the freeware CoDaPack. Math. Geol. 37(7): 773-793.

4th International Workshop on Compositional Data Analysis
May 9-13, 2011 Sant Feliu de Guíxols, Girona, Spain