



HAL
open science

Compositional discriminant analysis through calibrated evidence functions

Paul-Gauthier Noé, Andreas Nautsch, Driss Matrouf, Pierre-Michel Bousquet, Jean-François Bonastre Overview

► **To cite this version:**

Paul-Gauthier Noé, Andreas Nautsch, Driss Matrouf, Pierre-Michel Bousquet, Jean-François Bonastre Overview. Compositional discriminant analysis through calibrated evidence functions. The 10th International Workshop on Compositional Data Analysis (CoDaWork2024), Jun 2024, Girona, Catalonia, Spain. hal-04625713

HAL Id: hal-04625713

<https://hal.science/hal-04625713v1>

Submitted on 26 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Overview

Compositional discriminant analysis is a non-linear discriminant function where the discriminant components form a calibrated isometric-log-ratio transformed likelihood function over the set of classes. The latter can be called *evidence function* and can be seen as a multiclass extension of the weight-of-evidence in Bayesian updating. The compositional nature of the approach is on the treatment of the likelihood function (not on the modelled data) allowing a novel analysis method.

Bayesian updating on the Aitchison simplex [1]

Considering a set of classes $\mathcal{C} = \{c_i\}_{1 \leq i \leq N}$ and an individual who wants to infer the class of an observation x , its posterior probabilities are given by:

$$P(x) = w(x) \oplus \pi, \iff \text{ilr } P(x) = \underbrace{\text{ilr } w(x)}_{\text{Evidence function}} + \text{ilr } \pi,$$

where $P(x) = [P(c_i | x)]_{1 \leq i \leq N}$, $w(x) = [P(x | c_i)]_{1 \leq i \leq N}$ and $\pi = [P(c_i)]_{1 \leq i \leq N}$ are respectively the posterior distribution, the likelihood function and the prior.

The idempotence and the distribution of evidence function

An evidence function $\mathbf{l}(x) = \text{ilr } w(x)$ is calibrated if it results in the same posterior probabilities whether $\mathbf{l}(x)$ or the data x is given [2]:

$$\forall i \in \llbracket 1, N \rrbracket, P(c_i | x) = P(c_i | \mathbf{l}(x)) \iff \mathbf{l}(\mathbf{l}) = \mathbf{l}(x) \text{ (idempotence),}$$

leading to the following proposition extending a well-known result for LLRs [3],

Proposition 1 [2]: If $\mathbf{l} | c_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$, then $\forall i \in \llbracket 2, N \rrbracket, \mathbf{l} | c_i \sim \mathcal{N}(\boldsymbol{\mu}_i(\boldsymbol{\Sigma}), \boldsymbol{\Sigma})$,

$$\text{where } \boldsymbol{\mu}_1 = \boldsymbol{\mu}_1(\boldsymbol{\Sigma}) = \mathbf{A}^{-1} \mathbf{B} \text{vec}(\boldsymbol{\Sigma}), \text{ and } \boldsymbol{\mu}_i(\boldsymbol{\Sigma}) = \boldsymbol{\mu}_1(\boldsymbol{\Sigma}) - \boldsymbol{\Sigma} \mathbf{a}_{i-1} - \sum_{j=1}^{i-2} \frac{1}{j+1} \boldsymbol{\Sigma} \mathbf{a}_j,$$

$\mathbf{A} \in \mathcal{M}_{N-1, N-1}(\mathbb{R})$ and $\mathbf{B} \in \mathcal{M}_{N-1, (N-1)^2}(\mathbb{R})$ are constant matrices, and $\mathbf{a}_i = \sqrt{\frac{i+1}{i}} \mathbf{e}_i$ where \mathbf{e}_i is the i^{th} vector of the standard canonical basis of \mathbb{R}^{N-1} .

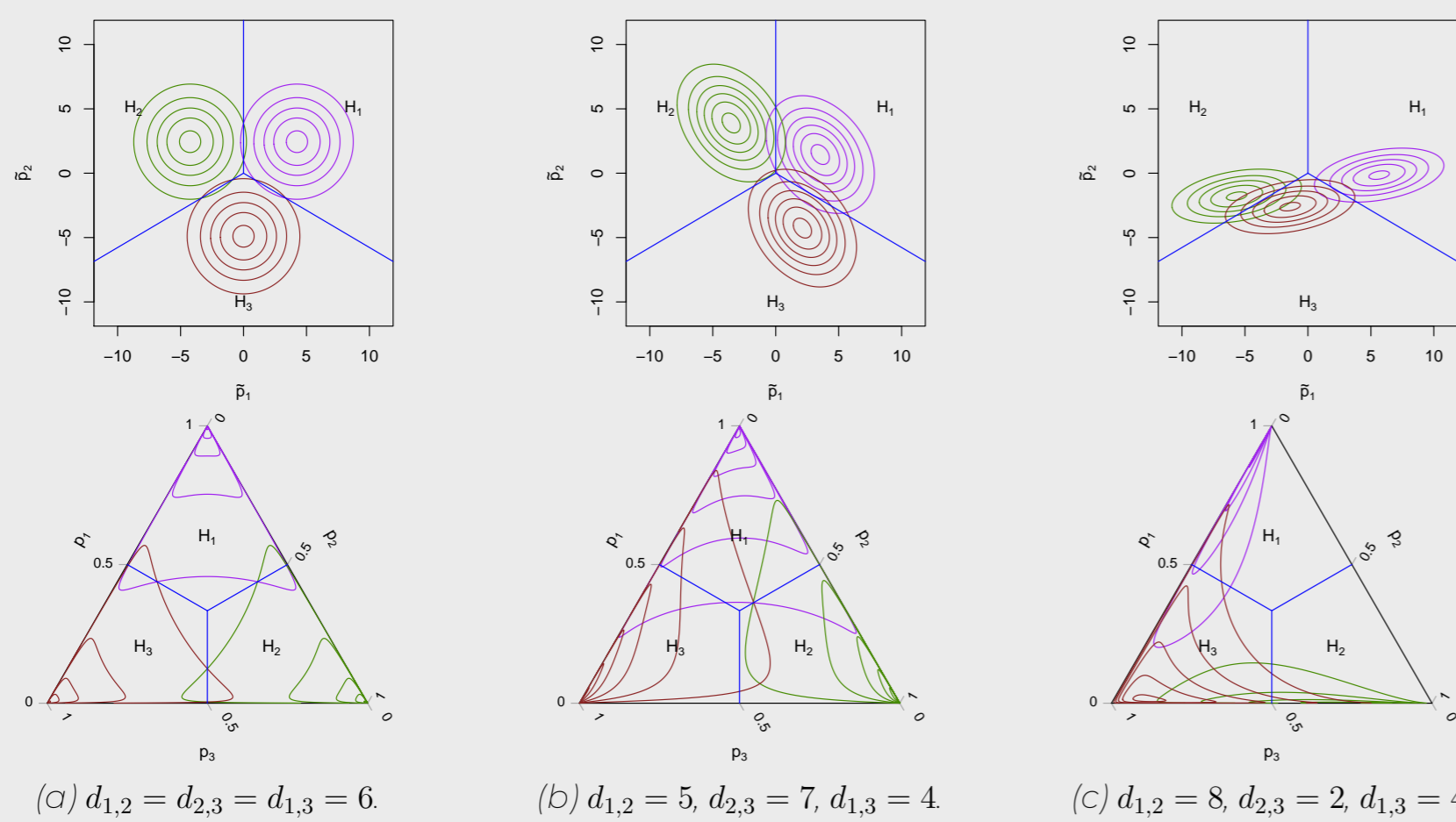


Figure: Gaussian conditional densities of the likelihood function in a 3 classes case. The covariance matrix can be expressed in terms of the three divergences between each density. The densities on the ILR space (top) are with respect to the Lebesgue measure while the densities on the simplex (bottom) are with respect to the Aitchison measure [4].

Toy experiments

With Gaussians

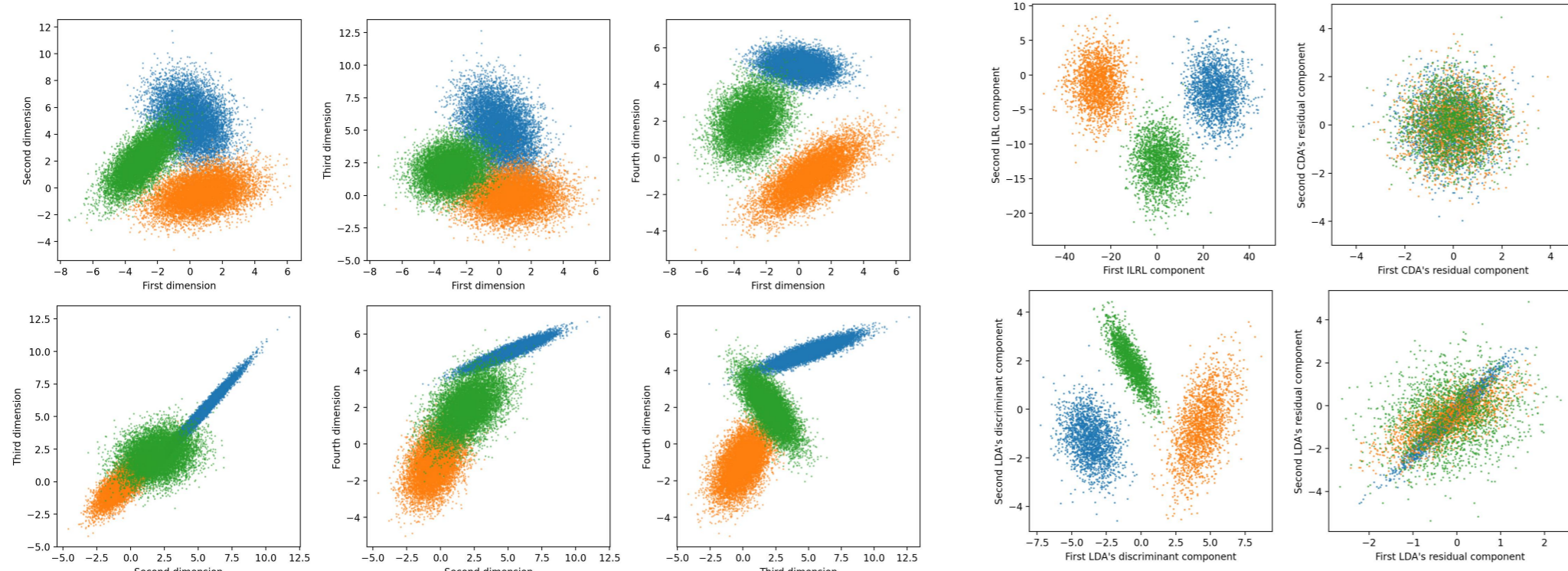


Figure: Training set.

Figure: Testing set in the CDA's (top) and LDA's (bottom) base space.

Table: C_{llr} measures [6]. Samples from the non-concerned class are discarded.

compared classes	LDA		QDA		CDA	
	C_{llr} [bit]	$C_{\text{llr}}^{\text{min}}$ [bit]	C_{llr}	$C_{\text{llr}}^{\text{min}}$	C_{llr}	$C_{\text{llr}}^{\text{min}}$
1 vs 2	$1.72 \cdot 10^{-3}$	0.0	0.0	0.0	$4.85 \cdot 10^{-5}$	0.0
1 vs 3	1.98	$1.43 \cdot 10^{-1}$	$1.72 \cdot 10^{-9}$	0.0	$9.46 \cdot 10^{-3}$	$5.04 \cdot 10^{-3}$
2 vs 3	$2.00 \cdot 10^{-1}$	$1.76 \cdot 10^{-2}$	$6.37 \cdot 10^{-4}$	0.0	$8.46 \cdot 10^{-3}$	$5.31 \cdot 10^{-3}$

With MNIST

0 1 2 3 4 5 6 7 8 9

The flattened images reduced to 40 principal components are used as inputs

Table: Cross-entropy and accuracy measures on the testing set for the MNIST's digit recognition task.

	empirical cross-entropy [nat]	accuracy [%]
LDA	$5.44 \cdot 10^{-1}$	87.67
QDA	$8.56 \cdot 10^{-1}$	96.24
CDA	$2.23 \cdot 10^{-1}$	94.43

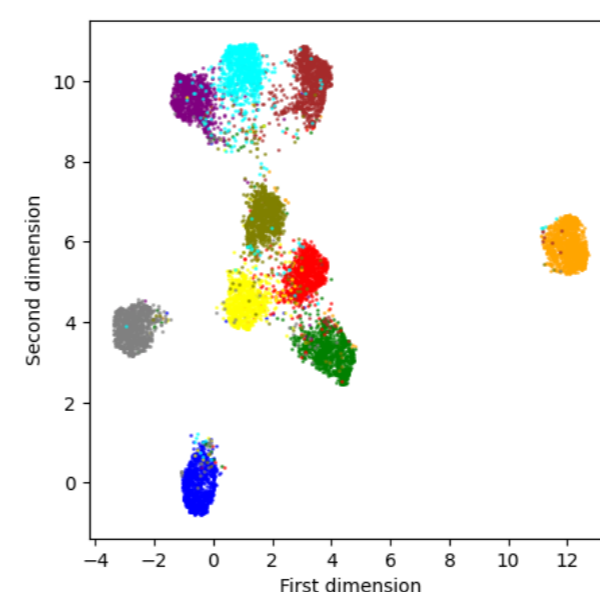


Figure: UMAP visualization of the evidence function.

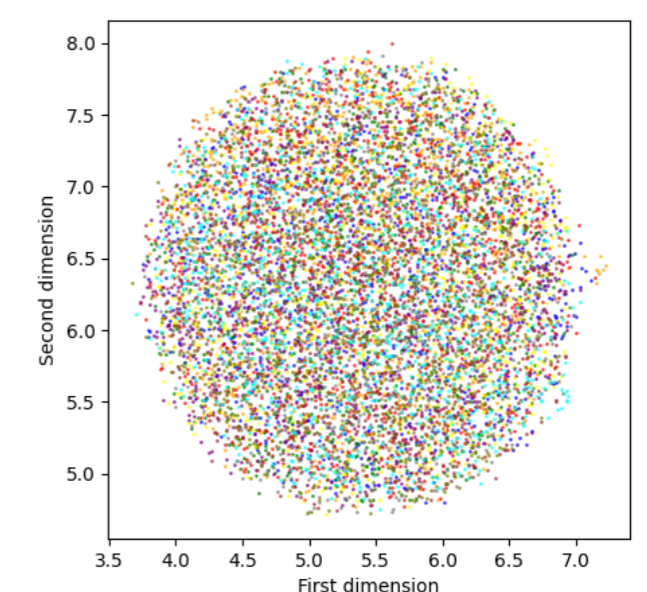


Figure: UMAP visualization of the residual.

Interpolation between digits in the space of evidence functions



Discussion

- The **discriminant components** form a calibrated isometric-log-ratio transformed likelihood function: an **evidence function**;
- No explicit assumption is made on the distribution of the data (except the existence of a diffeomorphism that would transform the data into the target Gaussians);
- Each discriminant component expresses a weight-of-evidence** in favour of a group of classes against another group according to the used Aitchison basis;
- This can be used for uncertainty-aware decisions avoiding under or overconfident predictions;
- In addition, the discriminant space has the **intuitive** Euclidean vector space structure given by the **Aitchison geometry**, allowing interpretability, distance measures, interpolations, and straightforward computation of the posterior probability distribution by simply shifting the evidence function by the prior.

References

- Juan José Egozcue and Vera Pawlowsky-Glahn. Evidence information in bayesian updating, 2011.
- Paul-Gauthier Noé. Representing evidence for attribute privacy: bayesian updating, compositional evidence and calibration, 2023.
- I.J. Good. Weight of evidence: A brief survey, 1985.
- Clòria Mateu-Figueras, Vera Pawlowsky-Glahn, and Juan José Egozcue. The principle of working on coordinates, 2011.
- Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp, 2017.
- Niko Brümmner and Johan du Preez. Application-independent evaluation of speaker detection, 2006.