Information, orthogonality and redundancy in Bayes updating

J. J. Egozcue\textsuperscript{(1)} and V. Pawlowsky-Glahn\textsuperscript{(2)}

\textsuperscript{(1)} Dept. Matemática Aplicada III
Universidad Politécnica de Cataluña, Barcelona, Spain
juan.jose.egozcue@upc.edu

\textsuperscript{(2)} Dept. Informática, Matemática Aplicada y Estadística
Universidad de Girona, Spain
vera.pawlowsky@udg.edu

Workshop Métodos Bayesianos 2014
Universidad Complutense, Facultad de Matemáticas, Madrid
November, 6-7, 2014
Summary

1. Bayes in the simplex
   - example
   - compositions
   - Bayes’ formula
   - joint and marginal information

2. Evidence-information
   - definition
   - properties

3. Redundancy
   - statement
   - procedure

4. Conclusions
The **geometry of the simplex** is an adequate framework to deal with discrete **Bayesian** updating.

- The **Aitchison norm** is a suitable measure of **information**

When joint information coming from **dependent sources** is not available:

- Use **marginal information** to update prior knowledge causes redundancy
- Bounds of **information/redundancy**
Spatial prediction

Gaussian random field: given variogram. Values of the field classified into 3 categories. Three dependent observed values. Goal: predict probabilities of categories at the unobserved position.

Example similar to Polyakova and Journel (2007)
**Compositions**

compositions in the simplex $S^n$

- vectors of $n$ positive components are compositionally-equivalent if they are proportional
- equivalence classes of compositionally-equivalent vectors are compositions
- compositions can be represented in the unit simplex $S^n$: vectors of constant sum

**examples**

- vectors of discrete probabilities: prior and posterior
- discrete likelihood: non-normalised

**generalisation to continuous cases:**
Bayes spaces: vector and Hilbert structure
Egozcue, Díaz-Barrero, Pawlowsky-Glahn (2006)
vanden Boogaart, Egozcue, Pawlowsky-Glahn (2010)
Egozcue, Pawlowsky-Glahn, Tolosana-Delgado, Ortego, vanden Boogaart (2013)
vanden Boogaart, Egozcue, Pawlowsky-Glahn (2014)
Aitchison geometry for compositions

The $n$-part simplex $S^n$ is an $(n - 1)$-dim Euclidean space vector space operations

- **group operation:** perturbation
  \[ x \oplus y = (x_1, \ldots, x_n) \oplus (y_1, \ldots, y_n) = C(x_1y_1, \ldots, x_ny_n) \]

- **multiplication by scalars:** powering
  \[ \alpha \odot x = C(x_1^{\alpha}, \ldots, x_n^{\alpha}) \]

**Euclidean metrics**

- **inner product:**
  \[ \langle x, y \rangle_a = \sum_{i>j} \log \frac{x_i}{x_j} \log \frac{y_i}{y_j} \]

- **norm, distance:**
  \[ \|x\|_a^2 = \langle x, x \rangle_a, \quad d_a(x, y) = \|x \ominus y\|_a \]
Bayes’ formula in the simplex

A discrete **parameter** to be estimated

R **observed** result of experiment (discrete)

**Bayes’ formula**

\[
P[A = i | R = r] = C \cdot P[R = r | A = i] \cdot P[A = i], \quad i = 1, 2, \ldots, n
\]

**Prior:** \[p(A) = (P[A = 1], \ldots, P[A = n])\]

**Likelihood:** \[q(r | A) = C(P[R = r | A = 1], \ldots, P[R = r | A = n])\]

**Posterior:** \[p(A | r) = (P[A = 1 | R = r], \ldots, P[A = n | R = r])\]

**Formulation in the** \(n\)-**part simplex** \(S^n\)

\[p(A | r) = q(r | A) \oplus p(A)\]
Coordinate representation

Cartesian coordinates (ilr) are available in $S^n$

$$b_1 = \frac{1}{\sqrt{2}} \log \frac{p_1}{p_2}$$

$$b_2 = \frac{1}{\sqrt{6}} \log \frac{p_1 p_2}{p_3}$$
Multiple observations $R_1 = r_1, \ldots, R_m = r_m$

**Joint likelihood:**
$q(r_1, \ldots, r_m|A) = (\ldots, P[R_1 = r_1, \ldots, R_m = r_m|A_i], \ldots)$

**Marginal likelihood**
$q(r_j|A) = (\ldots, P[R_j = r_j|A_i], \ldots), \ j = 1, \ldots, m$

$m$ independent observations conditional to $A$

$q(r_1, \ldots, r_m|A) = q(r_1|A) \oplus q(r_2|A) \oplus \cdots \oplus q(r_m|A)$

$m$ dependent observations

$q(r_1, \ldots, r_m|A) \neq q(r_1|A) \oplus q(r_2|A) \oplus \cdots \oplus q(r_m|A)$
Redundancy for non-exchangeable observations

redundancy: $\bigoplus_{k=1}^{m} q(r_k|A) \ominus q(r_1, \ldots, r_m|A)$

$L: R_1 = 1$  
$L: R_2 = 1$

simplex coordinates Joint-L 113  
$L: R_3 = 3$
Evidence-information

**Evidence vectors in** $S^n$: prior $\pi$, likelihood $q$, posterior $p$

**Evidence information** ($I_e$) is the *norm* of an evidence vector

$$I_e(z) = \|z\|_a, \quad z = p, q, \pi \in S^n$$

**Consequences**
- null e-information: neutral element, $I_e(n) = 0$
- opposite evidences: $I_e(z) = I_e((-1) \odot z)$
- orthogonal evidences $q_1, q_2$:
  $$I_e^2(q_1) + I_e^2(q_2) = I_e^2(q_1 \oplus q_2)$$
- distance of evidences $q_1, q_2$:
  $$d_a(q_1, q_2)$$

Evidence and Shannon information

Comparison of e-information and Kullback-Leibler divergence to the neutral element (Shannon information) along two straight-lines in the simplex.

they are not equal but not very disimilar!
Properties of e-information

- **scale-invariant** (applicable to likelihood)
  \[ I_e(q) = I_e(Cq) \]

- Bayes formula is addition of evidence vectors
  \[ p = q \oplus \pi \]

- likelihood-evidence does not depend on prior-evidence
  \[ I_e(p_1 \ominus p_2) = I_e(\pi_1 \ominus \pi_2) = I_e(q) \]

- appending evidence vectors: \( p_i \) with \( n_i \) parts and
  \[ p = (a_1 p_1, a_2 p_2) \]
  \[ I_e^2(p) = I_e^2(p_1) + I_e^2(p_2) + I_e^2(a_1 g_{n_1}(p_1), a_2 g_{n_2}(p_2)) \]
Redundancy

Framework:

- joint likelihood of experiments, \( q(r_1, \ldots, r_m | A) \), not available
- marginal likelihood of non-exchangeable experiments \( q(r_j | A) \), \( j = 1, \ldots, m \), known

Redundancy: is a composition (evidence vector); its norm is redundant e-information

\[
q_{RED} = \bigoplus_{j=1}^{m} q(r_j | A) \ominus q(r_1, \ldots, r_m | A)
\]

Upper bound of e-information

\[
\left\| \bigoplus_{j=1}^{m} q(r_j | A) \right\|_a \leq \sum_{j=1}^{m} \left\| q(r_j | A) \right\|_a
\]
Retrieval of e-information from non-exchangeable observations

Strategies:
- **use the only the likelihood** \( \arg \max_j [\|q(r_j|A)\|_a] \)
  loosing possible non-redundant information
- **tau-method approach** (weighted simplicial linear combination)
  \[ q(r_1, \ldots, r_m|A) \approx (\tau_1 \circ q(r_1|A)) \oplus (\tau_2 \circ q(r_1|A)) \oplus \cdots \oplus (\tau_k \circ q(r_m|A)) \]
  estimation of \( \tau \)'s ????
- **try to retrieve surely non-redundant e-information**
  present approach based on...

Conjecture:
if \( q(r_k|A) \) and \( q(r_\ell|A) \) are Aitchison-orthogonal, then the two observations \( R_j \) and \( R_\ell \) are exchangeable

Intuitive principles: for consecutive observations
- a single observation is non-redundant
- orthogonal evidence is non-redundant
- opposite evidence cancels previous information
- elimination of possibly redundant evidence depends on the order of the observations

Procedure (ordered observations)
Example (2,3,3): non-redundant minimal e-information

- **Red**: joint likelihood
- **Green**: marginal likelihood; **bullet**: mean marginal e-information
- **Violet triangles**: non-redundant minimal information permutation dependent (triangles)
- **Orange bullet**: mean non-redundant minimal information
Example (1,1,3): non-redundant minimal e-information

- **Red**: joint likelihood
- **Green**: marginal likelihood; **bullet**: mean marginal e-information
- **Violet**: non-redundant minimal information (triangles) shown permutation $R_1, R_3, R_2$
- **Orange bullet**: mean non-redundant minimal information
Conclusions

- **compositional-simplicial representation** is an adequate framework for (discrete) Bayesian updating
  - likelihood, prior and posterior are evidence vectors in $S^n$
  - perturbation is the Bayes’ formula (group operation)

- **Aitchison norm is scalar evidence-information**
  - properties of a scalar information, particularly,
  - likelihood-evidence does not depend on prior-evidence

- **redundancy when using non-exchangeable observations**
  - intuitive procedure to retrieve non-redundant evidence
  - **further research** is needed to find rigorous retrieval procedures
References


