Bayes spaces: use of improper priors and distances between densities

J. J. Egozcue[1], **V. Pawlowsky-Glahn**[2],
R. Tolosana-Delgado[1], **M. I. Ortego**[1]
and G. van den Boogaart[3]

[1] **Universidad Politécnica de Cataluña**, Spain
[2] **Universidad de Girona**, Spain
[3] **T.U. Freiberg**, Germany

Workshop on Bayesian Methods, UCM
Madrid, November 11, 2011

Distances/divergences for probability densities

applications:

- goodness of fit
- fitting distributions (e.g. kernel estimation)
- information theory (e.g. Kullback-Leibler div.)

examples of distances/divergences:

- from functional spaces: L^1, L^2, L^∞ applied to pdf's or cdf's
- ad-hoc: Hellinger-Matusita; Chi-square;
- from information theory: Kullback-Leibler, ...

What is lacking in these distances/divergences?

compatibility with probabilistic operations

two relevant operations:

- **convolution of pdf's**: associated with sum of random variables.
- **Bayes updating**: information acquisition

there is a need of a meaningful algebraic/geometric structure associated with Bayes updating

simplex and compositional data

composition: equivalent class of real vectors with proportional positive components

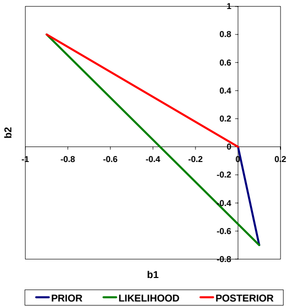
- components quantify parts of a whole
- only ratios between components are informative
- standard representative: a point in the simplex (components adding to 1)

Euclidean structure of the simplex

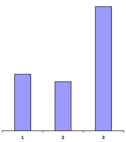
- interpretable operations: perturbation $\oplus$, powering $\odot$
- Aitchison metrics: inner product, norm and distance
- orthogonal bases, reference measure

perturbation in the simplex is the Bayes formula for discrete probability vectors

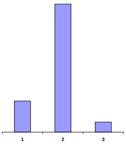
Coordinate representation



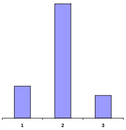
simplex coordinates



prior



likelihood



posterior

balance-coordinates

$$b_1 = \frac{1}{\sqrt{2}} \log \frac{p_1}{p_2} \qquad b_2 = \sqrt{\frac{2}{3}} \log \frac{(p_1 p_2)^{1/2}}{p_3}$$

leading ideas and goal

heuristic idea: a histogram is a composition and it is equivalent to a simplex element

- increasing the number of classes in a histogram it approaches a pdf...
- perturbation in the simplex is discrete Bayes formula, it can be extended to the continuous case...

goal

- vector space structure with Bayes updating as addition
- metric spaces of densities
- Hilbert spaces of densities

λ-equivalent measures

assuring existence of densities

measurable space: $(\Omega, \mathfrak{A})$

sigma-additive measures: λ equivalent to μ (finite or infinite)

$$\mu \equiv \lambda \ \Leftrightarrow \forall A \in \mathfrak{A}, (\ \lambda(A) = 0 \Leftrightarrow \mu(A) = 0\ )$$

- λ and μ have the same support
- the Radon-Nikodým derivative (density) exists

$$\frac{d\mu}{d\lambda} = f_\mu$$

Examples of equivalent measures:

- $\Omega = \mathbb{R}$: normal, t-student, Lebesgue measure (improper uniform)
- $\Omega = \mathbb{R}_+$: log-normal, gamma, Lebesgue measure in $\mathbb{R}_+$
- $\Omega = \{0, 1, 2, 3, \dots\}$: Poisson, geometric, counting measure

B-equivalence: proportional densities

reference measure λ**:** $\mu_1, \mu_2 \equiv \lambda$
densities: $f_1 = d\mu_1/d\lambda$, $f_2 = d\mu_2/d\lambda$

μ_1,μ_2 are B-equivalent, $\mu_1 =_B \mu_2$, iff

$$\exists c > 0,\ \forall A \in \mathfrak{A},\ \mu_1(A) = c \cdot \mu_2(A),\ f_1 = c \cdot f_2,\ (\lambda - a.e.)$$

Remarks

- **likelihood principle:** proportional likelihood functions convey identical information
- **normalization of probabilities:** not essential
- **essential information:** ratios of probabilities

perturbation and powering

Bayes space, reference λ: elements of $B(\lambda)$ are classes of B-equivalent measures/densities

perturbation (addition, group operation): $f_1, f_2 \in B(\lambda)$

$$f_1 \oplus f_2 =_B f_1 \cdot f_2 \ (\lambda - a.e.) \quad , \quad (\mu_1 \oplus \mu_2)(A) = \int_A \frac{d\mu_1}{d\lambda} \frac{d\mu_2}{d\lambda} \, d\lambda$$

powering (multiplication): $f \in B(\lambda)$, $\alpha \in \mathbb{R}$

$$\alpha \odot f =_B f^\alpha \ (\lambda - a.e.) \quad , \quad (\alpha \odot \mu)(A) = \int_A \left(\frac{d\mu}{d\lambda} \right)^\alpha \, d\lambda$$

vector space and Bayes theorem

$B(\lambda)$-space includes

- **prior densities**, proper or improper
- **likelihood functions**, integrable or not
- **posterior densities**, proper or improper

Bayes theorem

$$p =_B L \oplus \pi =_B \left(\bigoplus_{i=1}^{n} L_i \right) \oplus \pi$$

perturbation/Bayes updating is an internal operation in $B(\lambda)$
powering means (linear) weighting
iterate perturbation: group properties allow improper intermediate steps

vector space and exponential families

exponential families, k-parametric, natural parameters

$$f(x|\vec{\theta}) = C(\vec{\theta})g(x)\exp\left[\sum_{j=1}^{k}\theta_j T_j(x)\right]$$

$B(\lambda)$ **expression**

$$f_\lambda(x) =_B g_\lambda(x) \oplus \bigoplus_{j=1}^{k}(\theta_j \odot \exp[T_j(x)])$$

k**-dimensional affine subspace**

- $g_\lambda(x)$ origen of the affine subspace
- $\exp[T_j(x)]$ basis of the subspace
- θ_j coordinates of $f_\lambda(x)$

probability densities are a convex cone of the k-dimensional affine subspace

example: distribution of a sample maximum

n-sample, distribution F, density f
density of sample maximum

$$f_M(x) = nf(x) \cdot [F(x)]^{n-1}$$

reference f_0

$$f_M(x) =_B \underbrace{\frac{f(x)}{f_0(x)}}_{origin} \oplus (n-1) \odot \underbrace{[F(x)]}_{direction}$$

the family is 1-parametric and follows a straight-line with n

centered log-ratio mapping

clr for compositions

$$clr(\vec{x}) = \log(x_1, x_2, \ldots, x_k) - \frac{1}{k}\sum_{j=1}^{k} \log x_j \quad , \sum_{j=1}^{k} clr_j(\vec{x}) = 0$$

clr in $B(P)$ reference P probability measure; density f_P
definition of clr, $f \in B(P)$

$$clr(f) = \log(f) - \frac{1}{P(\Omega)} \int \log(f(x))\, f_P(x)\, dx$$

P prob. measure $\Leftrightarrow P(\Omega) = 1$

clr mapping is linear; scale and $B(P)$-reference invariant

$B^q(P)$ spaces

$B^q(P)$ space of measures/densities, $1 \leq q < \infty$

$$B^q(P) = \left\{ f \in B(P) : \int |\log f(x)|^q \, f_P(x) dx < +\infty \right\} ,$$

- clr exists for densities in $B^1(P)$;
- $clr : B^1(P) \to L^1_0(P)$ is one-to-one
- $B^1(P) \supseteq B^2(P) \supseteq \cdots \supseteq B^\infty(P)$
- $B^q(P)$ are Minkowsky metric spaces

$$d_{B^q}(f_1, f_2) = d_{L^q}(clr(f_1), clr(f_2)) = \left[\int (clr(f_1) - clr(f_2))^q \; dP \right]^{1/q}$$

B-derivative

$f(x|t) \in B^1(P)$; t external variable (time, space, sample values)
$f : \mathbb{R} \to B^1(P)$

Definition of B-derivative

$$\frac{d^{\oplus}}{dt} f(x|t) =_B \lim_{h \to 0} \frac{1}{h} \odot [f(x|t+h) \ominus f(x|t)]$$

if it exists. $\ominus = \oplus(-1)\odot$

- describes change of densities with t
- differential calculus and differential equations for densities/measures
- useful concept in applications (Bayesian, robust stats.)

Hilbert space

$B^2(P)$ is a separable Hilbert space

$$clr : B^2(P) \leftrightarrow L_0^2(P)$$

inner product, $f_1,\, f_2 \in B^2(P)$

$$\langle f_1, f_2 \rangle_{B^2} = \langle clr(f_1), clr(f_2) \rangle_{L^2}$$

distance and norm

$$d_{B^2}(f_1, f_2) = d_{L^2}(clr(f_1), clr(f_2)) \quad , \quad \|f_1\|_{B^2} = \|clr(f_1)\|_{L^2}$$

Hilbert basis and Fourier coordinates

$\psi_0, \psi_1, \psi_2, \ldots$ a Hilbert basis in $L^2(P)$
$\psi_0(x)$ constant function
Hilbert basis of $B^2(P)$

$$\exp(\psi_1), \exp(\psi_2), \ldots$$

coordinates: Fourier coefficients, $f \in B^2(P)$

$$f =_B \bigoplus_{j=1}^{\infty} \langle f, \exp(\psi_j) \rangle_{B^2} \odot \exp(\psi_j)$$

- Fourier coefficients are real orthogonal coordinates
- if normalized, distances, norms, orthogonal projections, ... are computed as ℓ^2 sequences
- they allow to use standard "real multivariate statistics"

the Normal family

reference $P = N(0,1)$ with Lebesgue density f_0
***P*-density** g corresponding to $N(m, \sigma^2)$

$$g(x) =_B f(x)/f_0(x) =_B \exp\left(-\frac{(x-m)^2 - \sigma^2 x^2}{2\sigma^2}\right)$$

clr

$$clr(g)(x) = \frac{x^2 - 1}{2} + \frac{1 - x^2 + 2mx}{2\sigma^2}$$

distance, $g_i \sim N(m_i, \sigma_i^2)$

$$d_{B^2}^2(g_1, g_2) = \frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right)^2 + \left(\frac{m_1}{\sigma_1^2} - \frac{m_2}{\sigma_2^2}\right)^2$$

Fourier expansion of Normal family

orthonormal Hilbert basis in $L^2(N(0,1))$: Hermite

$$\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} H_j(2^{-1/2}x)H_k(2^{-1/2}x)\, e^{-x^2/2}\, dx = \delta_{jk}K_j^{-2}\ ,\ K_j = 2^{-j/2}(j!)^{-1/2}$$

Hilbert basis in $B^2(N(0,1))$

$$\exp[\psi_j(x)] = \exp[K_j H_j(2^{-1/2}x)],\ j = 1, 2, \ldots$$

Fourier expansion

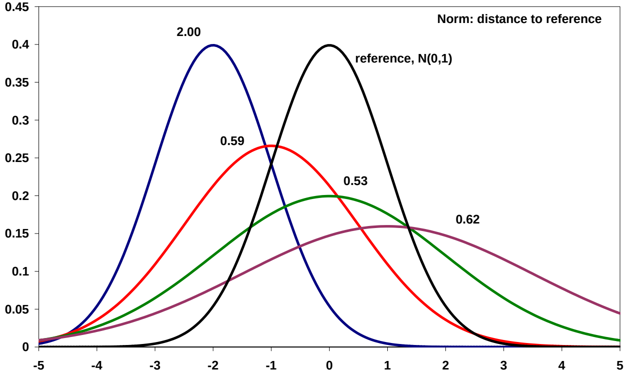
$$g(x) =_B c_1 \odot \exp[\psi_1(x)] \oplus c_2 \odot \exp[\psi_2(x)]$$

$$c_1 = \frac{m}{\sigma^2}\ ,\ c_2 = -\sqrt{2}\left(\frac{1}{2\sigma^2} - \frac{1}{2}\right)\ ,\ c_j = 0,\ j = 3, 4, \ldots,$$

norms of normals

reference: $N(0,1)$

$N(-2,1)$, $N(-1,1.5^2)$, $N(0,2^2)$, $N(1,2.5^2)$



conclusions

results

- **proportional densities are considered equivalent ($B(\lambda)$)**
  - perturbation is the (extended) Bayes updating
  - proper and improper priors, likelihoods and posteriors are in $B(\lambda)$
  - $(B(\lambda), \oplus, \odot)$ is a vector space
  - linear affine subspaces contain exponential families
- **q-log-integrable densities ($B^q(P)$) are metric spaces**
  - $clr : B^1(P) \to L_0^1(P)$ is one-to-one (isometry)
- **2-log-integrable densities in $B^2(P)$ are a separable Hilbert space. Standard tools are then available:**
  - Hilbert basis and Fourier expansions
  - distances, norm, orthogonal projections
  - Aitchison geometry of the simplex is a particular case.

conclusions

consequences

- the new framework allows to rephrase most standard probabilistic models (Bayes theorem, exponential families, ...) in a simple and formal way
- tools of vector, metric and Hilbert spaces are now available for probabilistic/statistical modelling

a good deal of research is still pending...

- the role of references in $B^2(P)$
- possible uses of Fourier transforms in $B^1(P)$ and $B^2(P)$
- asymptotic theory on $B(\lambda)$
- characteristics of well-known families (normal, gamma, beta, t-student,...)
- approximation of $B^2(P)$ spaces by the simplex geometry

some steps towards B-spaces

Aitchison, J. (1986). The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). 416 p.

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

Egozcue, J. J., J. L. Díaz-Barrero and V. Pawlowsky-Glahn (2006). Hilbert Space of Probability Density Functions Based on Aitchison Geometry, Acta Mathematica Sinica 22 (4), 1175–1182

van den Boogaart, K. G., J. J. Egozcue, and V. Pawlowsky-Glahn (2010). Bayes linear spaces. Statistics and Operations Research Transactions, SORT 34 (2), 201–222

example: zero-inflated Poisson exponential family

reference measure: counting measure

$\nu(x) = 1, x = 0, 1, 2, \ldots$

mixture expression

$$f(x|\phi, p) = (1-p)\cdot\delta(x) + p\cdot\frac{\phi^x e^{-\phi}}{x!}$$

$B(\nu)$ **2-parametric exponential family**

$$f(x|\theta_1, \theta_2) =_{B(\nu)} \underbrace{\frac{1}{x!}}_{origin} \oplus \left(\theta_1 \odot \underbrace{e^x}_{basis_1}\right) \oplus \left(\theta_2 \odot \underbrace{e^{\delta(x)}}_{basis_2}\right)$$

$$\theta_1 = \log\phi\ ,\ \theta_2 = \log\left[(1-p)e^{\phi} + p\right], C(\theta_1, \theta_2) = [\exp(\theta_2) + \exp(\exp(\theta_1)) - 1]^{-1}$$

3-parametric conjugate family

$$\pi(\theta_1, \theta_2) =_B \exp\left(t_0 \log C(\theta_1, \theta_2) + t_1\theta_1 + t_2\theta_2\right)$$