A forgotten space: the parameter space

J. J. Egozcue[1], **V. Pawlowsky-Glahn**[2] and **M. I. Ortego**[1]

[1]Dep. Civil and Environmental Eng., U. Politécnica de Cataluña, Barcelona, Spain
[2]Dep. Inf., Mat. Apl. y Estadística; Uni. Girona, Girona, Spain

Métodos Bayesianos'17.
Madrid, November 7-8, 2017

Sample space

Definition:
Given a random object X, the sample space $\mathcal{S}$ of X is

- a set containing all possible values of X
- with a sigma-field of events where probability is to be evaluated (minimal structure)

However, if the analysis requires operations (averages, distances, projections, ...) they need to be defined in $\mathcal{S}$

the sample space is more than a set

its structure determines the available tools for the analysis

Examples of sample space: random sets

Random non-empty compact sets in a plane

- distance between sets: Hausdorff
  $d_H(S_1, S_2) = \max\{\sup_{s_1 \in S_1} d(s_1, S_2), \sup_{s_2 \in S_2} d(S_1, s_2)\}$
- sigma-field generated by Hausdorff open balls
- operation: set union, intersection
- sample variability: $\mathrm{Var}[X; S] = \sum d_H^2(X_i, S)$
- sample mean and variance (Fréchet approach)
  $Mean(X) = \mathrm{argmin}_S \mathrm{Var}[X; S] \quad , \quad \mathrm{Var}(X) = \min \mathrm{Var}[X; S]$
- ...

Examples of sample space: time to failure T

Two alternatives

- sample space: $\mathbb{R}_+$
- group operation: $x \oplus y = \exp(\log x + \log y)$
- square distance: $d_+^2(x,y) = \log^2(x/y)$
- mean value: $\exp[(\sum \log x_i)/n]$ (geometric mean)
- reference measure: $\lambda_+\{(a,b)\} = |\log a/b|$

- sample space: $\mathbb{R}$
- group operation: $x + y$
- square distance: $d^2(x,y) = (x-y)^2$
- mean value: $\sum x_i/n$ (arithmetic mean)
- reference measure: Lebesgue measure $\lambda\{(a,b)\} = |b-a|$

Examples of sample space: random densities

Random positive densities supported on an interval

- **Bayes Space:** elements are equivalence classes of proportional densities
- **group operation:** perturbation $f_1(x) \oplus f_2(x) = Cf_1(x)f_2(x)$
- **transformation:** $clr(f(x)) = \log f(x) - (1/L) \int log(f(x))$
- **inner product:** $\langle f_1, f_2 \rangle_a = \langle clr(f_1), clr(f_2) \rangle$
- **mean value:** geometric mean
- **coordinates:** Fourier coefficients
- **structure:** **Hilbert space**

Bayesian statistics: what is there?

There are: random observations, random parameters jointly distributed and the corresponding sample spaces (reference measures assumed Lebesgue, λ)

- **Initial/prior and final/posterior probability measures:** $P_0\{A\}$, $P_1\{A|R\}$ and Lebesgue densities $f_0(\theta)$, $f_1(\theta|R)$
- **likelihood conditional to observation** R**,** $L(\theta|R)$

Bayes theorem: For any event A in the **parameter (sample) space**

$$f_1(\theta|R) = C \cdot L(\theta|R) \cdot f_0(\theta) \quad , \quad P_1\{A|R\} = \int_A f_1(\theta|R) d\lambda$$

observation and parameter sample spaces should be specified!

reference measures determine which **densities** and their characteristics are to be used

Binomial observations with two different references

Observations: $X = 1$ successes, $N = 10$ trials
Parameter: $p \in (0, 1)$
Initial probability: $Beta(a = 1/2, b = 1/2)$

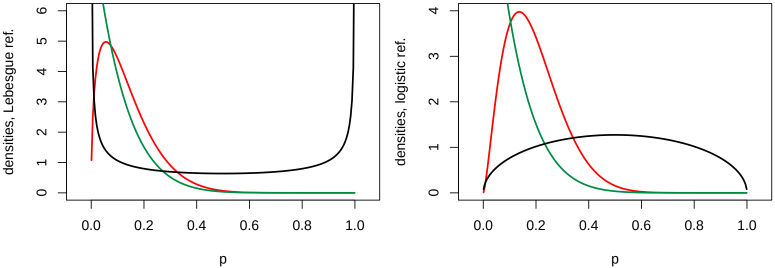
Lebesgue reference	**Logistic reference**
uniform measure in (0,1)	$\propto [p(1-p)]^{-1}$
$f_0(p) \propto p^{a-1}(1-p)^{b-1}$	$f_0(p) \propto p^{a}(1-p)^{b}$

$$L(p|X = 1, N = 10) \propto p^X(1-p)^{N-X}$$

$f_1(p) \propto p^{a+X-1}(1-p)^{b+N-X-1}$	$f_1(p) \propto p^{a+X}(1-p)^{b+N-X}$

Binomial observations with two different references

Likelihood, Initial, Final



Same sample set; different reference measure

Lebesgue measure λ, uniform in $(0, 1)$

Logistic measure μ, $d\lambda/d\mu = p(1-p)$

Densities are Radom-Nikodym derivatives

$$f^{(Leb)}(p) = \frac{dP}{d\lambda} \qquad f^{(Logis)}(p) = \frac{dP}{d\lambda}\frac{d\lambda}{d\mu}$$

What is changed by different references?

Almost nothing:

Densities are different but probability measures are equal.
However, modes of final densities are not equal
Not a problem for "Bayesians" (hope so) but for "frequentists"

...

Is maximum likelihood flawed?

- Yes, unless the reference measure (Lebesgue) is specified for likelihood and posterior densities;
- Does it mean that parameters only can live in $\mathbb{R}$?
- What about (max) likelihood ratio tests?

Binomial example: maximum posterior point estimates of p are

reference	initial dist.	initial mode	final mode
Lebesgue	Beta(1,1) (unif)	undefined	$1/10 = 0.10$
Lebesgue	Beta(0.5,0.5) (Jeffreys)	0 and 1	$1.5/10.5 = 0.143$
Logistic	Beta(0.5,0.5) (Jeffreys)	0.50	$2.5/11.5 = 0.217$

Number of extreme events in time

Observation: number of events per year

$N(t) \sim \text{Poisson}(\theta(t)) \quad , \quad \theta(t) = a + bt + cI\{t > t_0\}$

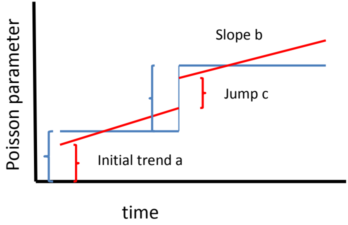
Parameters:

Initial Poisson parameter: $a > 0$ (positive, $\mathbb{R}_+$)

Slope (climatic change?): lower bound depends on a and c

Change (measuring device?): lower bound depends on a, b

Changing one of these parameters implies the change of the others to preserve fitting of data



Lack of orthogonality!

Can a, b and c be independent?

Which is the structure of the parameter space?

Final descriptors depend on the parameter space

- **Reference measure**
  $\Rightarrow$ **initial and final probability densities**
- **Vector space operations & distance**
  $\Rightarrow$ **mean and variance**
- **Inner product** $\Rightarrow$ **orthogonality, projections**

Mean and variance depend on distances (Fréchet)

$$Var[\Theta;\eta] = \int d^2(\theta;\eta)\ f(\theta)\ d\theta$$

$$Mean[\Theta] = argmin_\eta\, Var[\Theta;\eta]\ ,\ \ Var[\Theta] = min_\eta\, Var[\Theta;\eta]$$

Illustrative example:
Estimation of multinomial parameters or their coordinates

The 3-part simplex $\mathbb{S}^3$ as a parameter space

Compositions:

- Vectors of proportional positive components are compositionally equivalent
- A composition is an equivalence class
- Compositions can be represented in $\mathbb{S}^3$

$$\mathbb{S}^3 = \left\{ \mathbf{x} \in \mathbb{R}^3 \;\middle|\; x_i > 0 \,,\, \sum x_i = 1 \right\}$$

Vector space operations $\mathbf{x}, \mathbf{y} \in \mathbb{S}^3$, $\alpha \in \mathbb{R}$

perturbation **powering**

$$\mathbf{x} \oplus \mathbf{y} = \mathcal{C}(x_1 y_1, x_2 y_2, x_3 y_3) \quad , \quad \alpha \odot \mathbf{x} = \mathcal{C}(x_1^\alpha, x_2^\alpha, x_3^\alpha)$$

Euclidean structure

Centered log-ratio

$$clr(\mathbf{x}) = \left(\log \frac{x_1}{g_m(\mathbf{x})}, \log \frac{x_2}{g_m(\mathbf{x})}, \log \frac{x_3}{g_m(\mathbf{x})}\right) \quad , \quad \sum_{i=1}^{3} clr_i(\mathbf{x}) = 0$$

Aitchison's inner product, norm and distance

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \langle clr(\mathbf{x}), clr(\mathbf{y}) \rangle \quad , \quad \|\mathbf{x}\|_a = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_a}$$

$$d_a(\mathbf{x}, \mathbf{y}) = \sqrt{d(clr(\mathbf{x}), clr(\mathbf{y}))}$$

$\mathbb{S}^D$ has a $(D-1)$-dim Euclidean space structure

any composition can be represented by coordinates, particularly, **Cartesian coordinates**

Isometric log-ratio (ilr) coordinates

Contrast $(D, D-1)$-matrix:

$$V^\top V = I_{D-1} \quad , \quad VV^\top = I_D - (1/D)\mathbf{1}^\top\mathbf{1}$$

ilr/Cartesian coordinates

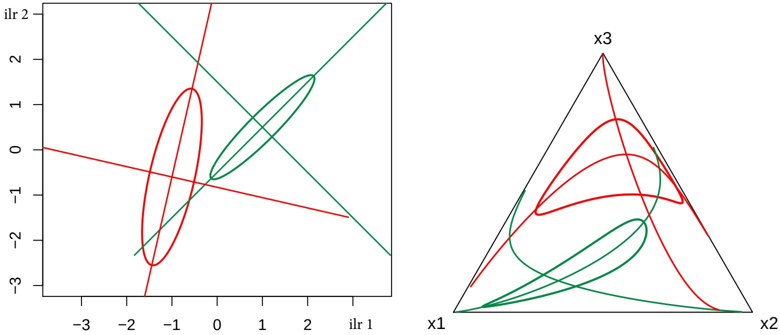
$$ilr(\mathbf{x}) = clr(\mathbf{x})\ V$$

Example for $\mathbb{S}^3$

$$b_1 = ilr_1(\mathbf{x}) = \sqrt{\frac{2}{3}} \log \frac{g_m(x_1, x_2)}{x_3} \quad , \quad b_2 = ilr_2(\mathbf{x}) = \sqrt{\frac{1}{2}} \log \frac{x_1}{x_2}$$

Compositional ellipses and axes

ilr-coordinates: straight-lines, angles, distances, convex domains
simplex: all is distorted; infinity is at the border . . .



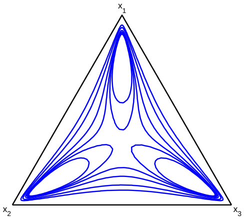
The Normal distribution in the simplex

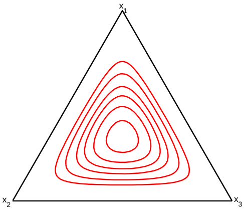
Aitchison measure in the simplex $\mathbb{S}^D$: λ_a
corresponds to Lebesgue measure in ilr-coordinates

$$B \quad \text{Borelian in } \mathbb{R}^{D-1} \quad , \lambda_a\{ilr^{-1}(B)\} = \lambda\{B\}$$

$$\frac{d\lambda}{d\lambda_a} = \sqrt{D}\, p_1 p_2 \cdots p_D$$

A random composition is normal in the simplex if its coordinates are multivariate normal





Densities of a Normal in the simplex

Reference measure: Lebesgue   Reference measure: Aitchison

Model of multinomial observations

Observations: counts in 3 categories from n trials,
$\mathbf{x} = (x_1, x_2, x_3)$,
Multinomial model:
parameters $\mathbf{p} = (p_1, p_2)$, $p_3 = 1 - p_1 - p_2$

$$Pr[\mathbf{X} = \mathbf{x} \mid \mathbf{p}] = \frac{n!}{\prod x_i!} \quad p_1^{x_1}\ p_2^{x_2}\ p_3^{x_3}$$

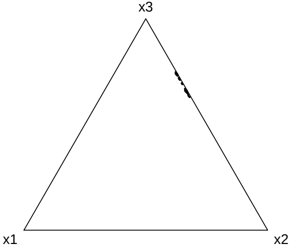
Naïve parameter space: $\mathbf{p}$ is in a 2-D affine subspace of $\mathbb{R}^3$

Compositional parameter space: $\mathbf{p}$ can be represented in the 3-part simplex or equivalently
represented by isometric-coordinates (ilr) in $\mathbb{R}^2$

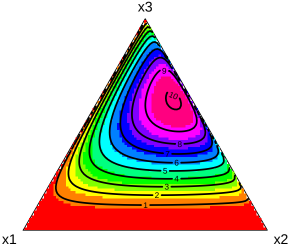
A Multinomial case

- observed counts: $x_1 = 0$, $x_2 = 35$, $x_3 = 70$
- Dirichlet initial: $\alpha = 0.3 \cdot (1, 25, 40)/66$
- simulated probabilities $\mathbf{p}^{(k)} = (p_{k1}, p_{k2}, p_{k3})$, k=1,2,. . . ,K
- or ilr-coordinates $\mathbf{b}^{(k)} = ilr(\mathbf{p}^{(k)}) = (b_{k1}, b_{k2})$,

Initial and final distribution of p in a ternary diagram



failed



expected but not real

A multinomial case: modes and means

Final distribution: **Dirichlet**$(0.0045,\ 35.11,\ 70.18)$
Comparisons between numbers should be relative!

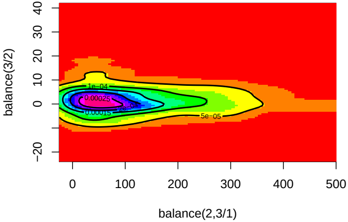
	ref. measure	p_1	p_2	p_3
mode	Lebesgue	0	0.330	0.670
mode	Aitchison	$4.32 \cdot 10^{-5}$	0.333	0.665
mean	Lebesgue	$4.32 \cdot 10^{-5}$	0.333	0.665
mean	Aitchison	$1.55 \cdot 10^{-98}$	0.332	0.668

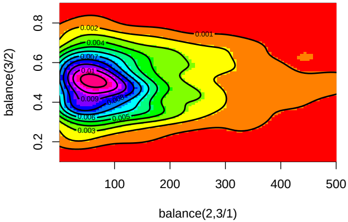
Formulas:

	ref. measure	formula
mode	Lebesgue	$(\alpha_i - 1)/(\sum \alpha_k - 3)$, $\alpha_i > 1$
mode	Aitchison	$\alpha_i / \sum \alpha_k$
mean	Lebesgue	$\alpha_i / \sum \alpha_k$
mean	Aitchison	$C \exp(\psi(\alpha_i))$

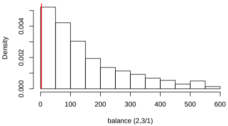
Multinomial case: ilr-coordinates

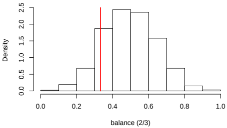
Initial and final distribution of b (note plot scale)





are prior Aitchison mean probabilities credible?



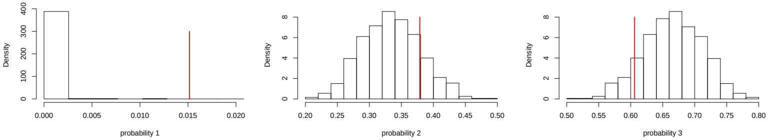


$Pr[b_1 \leq 2.82] = 0.035$ $Pr[b_2 \leq 0.33] = 0.134$

Multinomial case: p marginals?

Parameters: $\mathbf{p} = (p_1, p_2, p_3), \;\; p_1 + p_2 + p_3 = 1$
A marginal on p_1 is not relevant as it does not convey information on odds like p_1/p_2 or p_1/p_3

It may lead to confusing results: are the values $p_1 = 0.015$, $p_2 = 0.38$, $p_3 = 0.61$ credible?



$Pr[p_1 \leq 0.015] = 0.999 \quad Pr[p_2 \leq 0.38] = 0.826 \quad Pr[p_3 \leq 0.61] = 0.093$

Marginalisation in Bayesian statistics

Types of marginalisation

Bayes factors: odds for two models H_0, H_1

$$\frac{f(H_1|x)}{f(H_0|x)} = \frac{\int f(\theta|H_1)L(\theta|x)\, d\theta}{\int f(\theta|H_0)L(\theta|x)\, d\theta} \cdot \frac{Pr(H_0)}{Pr(H_1)}$$

Marginals

$$f(\theta_1|\mathbf{x}) = \int f(\theta_1, \theta_2|\mathbf{x})\, d\theta_2 = \int f(\theta_1|\theta_2, \mathbf{x}) \cdot f(\theta_2|\mathbf{x})\, d\theta_2$$

Expectations of predictive quantities

$$E[\varphi(\mathbf{y})|\mathbf{x}] = \int_\theta \varphi(\mathbf{y}) f(\mathbf{y}|\theta, \mathbf{x})\, d\theta$$

All these integrals can be viewed as expectations ...

Marginals or expected conditionals?

Generic expression of a marginal: (initial or final)

$$f(\theta_1) = \int_{\theta_2} f(\theta_1|\theta_2)\cdot f(\theta_2)\, d\theta_2 = E_{\theta_2}[f(\theta_1|\theta_2)] \quad , \quad \theta_1 \in \mathbb{R}^k, \theta_2 \in \mathbb{R}^\ell$$

Real setup for $f(\theta_1|\theta_2)$, in $\mathcal{L}(\mathbb{R}^k)$ If $f(\theta_1|\theta_2)$ is considered as a function in $\mathcal{L}^1(\mathbb{R}^k)$ then

$$f(\theta_1) = \int_{\theta_2\in\mathbb{R}^\ell} f(\theta_1|\theta_2)\cdot f(\theta_2)\, d\theta_2$$

is the traditional **probabilistic marginal of θ_1 and also its mean conditional**

Bayes space setup for $f(\theta_1|\theta_2)$

$$E_{\theta_2}[f(\theta_1|\theta_2)] = \exp\left(\int_{\theta_2\in\mathbb{R}^\ell} \log(f(\theta_1|\theta_2))\cdot f(\theta_2)\, d\theta_2\right)$$

which is the **mean conditional and the geometric marginal of θ_1**, and it is different from the probabilistic marginal

Conclusions

- Almost always parameters are considered to be real, even in cases in which this assumption is not appropriate
- If parameters are assumed to be in a subset of real space, initial and final distributions in a Bayesian framework are equal whatever the parameter space structure. However,
  - the mode of parameters depends on the assumed reference measure
  - means and variances depend on the distance assumed
  - other moments can depend as well ...
- Selecting appropriate real parameters may simplify computation and representation of initial and final distributions
- When densities are assumed to be in a space of functions different from $\mathcal{L}(\mathbb{R}^k)$, marginals and mean conditionals can differ

References

van den Boogaart, K. G., J. J. Egozcue, and V. Pawlowsky-Glahn, (2014): Bayes Hilbert spaces. Australian and New Zealand Journal of Statistics, 56(2), 171-194.

Egozcue, J. J., Barceló-Vidal, C., Martín-Fernández, J. A., Jarauta-Bragulat, E., Díaz-Barrero, J. L. and Mateu-Figueras, G. (2011): Elements of simplicial linear algebra and geometry. In Pawlowsky-Glahn, V. and Buccianti A. (Eds.) Compositional Data Analysis: Theory and Applications, Wiley, Chichester UK

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

Egozcue, J. J., V. Pawlowsky-Glahn, R. Tolosana-Delgado, M. I. Ortego, and K. G. van den Boogaart, (2013): Bayes spaces: use of improper distributions and exponential families. Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales, Serie A, Matemáticas (RACSAM), 107, 475-486

Fréchet, M. (1948): Les éléments Aléatoires de Nature Quelconque dans une Espace Distancié. Annales de l'Institut Henri Poincaré, 10 (4), 215-308

D. V. Lindley (1988): Statistical Inference concerning Hardy-Weinberg equilibrium. In Bayesian Statistics 3, Eds. Bernardo, J. M., DeGroot, M. H. and Lindley, D. V. and Smith, A. F. M., Oxford University Press, 307–326

G. Monti, Mateu-Figueras, G., and Pawlowsky-Glahn, V. (2011): Notes on the Scaled Dirichlet Distribution, John Wiley & Sons, Ltd, pp. 128–138

M. I. Ortego, J. J. Egozcue and R. Tolosana-Delgado (2014): Bayesian trend analysis of extreme wind using observed and hindcast series off the Catalan coast, NW Mediterranean Sea, Natural Hazards and Earth System Science, 14, 9, 2387–2397

V. Pawlowsky-Glahn, J. J. Egozcue and R. Tolosana-Delgado (2015): Modeling and Analysis of Compositional Data, John Wiley & Sons, Chichester, UK.