HILBERT SPACE OF PROBABILITY DENSITY FUNCTIONS WITH AITCHISON GEOMETRY

J. J. Egozcue[1] **and J. L. Díaz-Barrero** [2]
[1]Universitat Politècnica de Catalunya, Barcelona, Spain; *juan.jose.egozcue@upc.es*
[2]Universitat Politècnica de Catalunya, Barcelona, Spain.

1 Introduction

Compositional data analysis motivated the introduction of a complete Euclidean structure in the simplex of D parts. This was based on the early work of J. Aitchison (1986) and completed recently when Aitchinson distance in the simplex was associated with an inner product and orthonormal bases were identified (Aitchison and others, 2002; Egozcue and others, 2003). A partition of the support of a random variable generates a composition by assigning the probability of each interval to a part of the composition. One can imagine that the partition can be refined and the probability density would represent a kind of continuous composition of probabilities in a simplex of infinitely many parts. This intuitive idea would lead to a Hilbert-space of probability densities by generalizing the Aitchison geometry for compositions in the simplex into the set probability densities.

This approach is appealing due to several circumstances. We remark some of them:

- perturbation between probability densities is just the operation implied in the Bayes theorem, as remarked by J. Aitchison for discrete probabilities, and now generalized to continuously defined variables and parameters;
- the approach suggests a new distance between probability densities, based on the principles of the standard Aitchison distance for compositions;
- probability densities would be approached by orthogonal series and least squares approximations would be allowed;
- probability densities would be represented by their coefficients on some orthonormal basis. This would allow a proper definition of expectation of random probability densities as they are used in Bayesian statistics (predictive distributions).

Other points could be pointed out but they should be developed in future research.

The present aim is to give the basic definitions of perturbation, power transformation, Aitchison inner product, norm and distance when generalized to probability densities whose support is an interval. Several technical aspects are also proved; they are mainly related to the required completeness of the Hilbert space. Finally, three Hilbert bases (orthonormal) are defined using the standard theory of $L^2(-\ell,\ell)$, the Hilbert space of square-summable functions in the interval $(-\ell,\ell)$.

2 Algebraic structure of probability density functions on finite intervals

In what follows, we will deal with probability density functions on a finite interval, $(-\ell,\ell)$, $\ell>0$, without loss of generality. That is, we will consider real functions $f:(-\ell,\ell)\to\mathbb{R}$ such that: (i) $0\le f(x)$ and (ii) $\int_{-\ell}^{\ell} f(x)\,dx=1$. Additionally, for our initial results, we also restrict our attention to bounded densities, $0<m\le f(x)<M$. We will denote by $\mathcal{A}_\ell^2$ the set of such bounded pdf's.

The fundamental algebraic composition laws are the perturbation and power transformation that we define as

Definition 2.1 *Let $f, g \in \mathcal{A}_\ell^2$ be any two probability density functions on $(-\ell, \ell)$. We define its perturbation as the function $\oplus : \mathcal{A}_\ell^2 \times \mathcal{A}_\ell^2 \to \mathcal{A}_\ell^2$ given by*

$$f \oplus g = \frac{f(x)g(x)}{\displaystyle\int_{-\ell}^{\ell} f(\xi)g(\xi)\, d\xi} = \mathcal{C}(fg). \tag{1}$$

Definition 2.2 *Let $f \in \mathcal{A}_\ell^2$ be a probability density function on $(-\ell, \ell)$ and let α be a real number. We define the power transformation of f as the function $\otimes : \mathbb{R} \times \mathcal{A}_\ell^2 \to \mathcal{A}_\ell^2$ given by*

$$\alpha \otimes f = \frac{f^\alpha(x)}{\displaystyle\int_{-\ell}^{\ell} f^\alpha(\xi)\, d\xi} = \mathcal{C}(f^\alpha). \tag{2}$$

Now, we can state and prove the main result on the algebraic structure of the set $\mathcal{A}_\ell^2$.

Theorem 2.1 *The set of probability density functions with the perturbation and the power transformation, $(\mathcal{A}_\ell^2, \oplus, \otimes)$, is a vector space.*

Proof. We have to prove the following properties:

(a) Commutative group structure of $(\mathcal{A}_\ell^2, \oplus)$. For any $f, g, h \in \mathcal{A}_\ell^2$ we have that

- operation (1), $f \oplus g$, is closed in the set $\mathcal{A}_\ell^2$. In fact, the product $f(x)g(x)$ is bounded as $0 < m_f m_g \leq f(x)g(x) \leq M_f M_g$, being m_f, M_f, m_g, M_g the lower and upper bounds for f and g respectively and the product is integrable.
- Perturbation is commutative, $f \oplus g = g \oplus f$.
- Perturbation is associative, $(f \oplus g) \oplus h = f \oplus (g \oplus h)$.
- There exist a unique neutral element $e(x) = 1/2\ell$ such that $f \oplus e = e \oplus f = f$.
- To each f corresponds a unique $\overline{f} = f^{-1}$ such that $f \oplus \overline{f} = \overline{f} \oplus f = e$.

(b) Power transformation properties. For any $f, g \in \mathcal{A}_\ell^2$ and $\alpha, \beta \in \mathbb{R}$ we have that

- operation (2), $\alpha \otimes f$, is in $\mathcal{A}_\ell^2$ for any real α. Indeed, $0 < m^\alpha \leq f^\alpha \leq M^\alpha$ if $\alpha \geq 0$ and $0 < M^\alpha \leq f^\alpha \leq m^\alpha$ if $\alpha < 0$ and f^α is integrable.
- Power transformation is associative, $\alpha \otimes (\beta \otimes f) = (\alpha \cdot \beta) \otimes f$.
- Neutral element, $1 \otimes f = f$.
- Power transformation is distributive with respect perturbation, $\alpha \otimes (f \oplus g) = (\alpha \otimes f) \oplus (\alpha \otimes g)$.
- Power transformation is distributive with respect to scalar addition, $(\alpha + \beta) \otimes f = (\alpha \otimes f) \oplus (\beta \otimes f)$.

The preceding properties immediately follow from (1) and (2) and the proof is completed. □

3 Geometric structure of probability density functions on finite intervals

In order to structure $\mathcal{A}_\ell^2$ as a pre-Hilbert space, we define the inner product of bounded probability densities. The associated norm and distance are also defined. These definitions are directly inspired in the corresponding inner product, norm and distance in the simplex which constitute the main features of the so called Aitchison geometry (Aitchison and others, 2002; Egozcue and others, 2003). This is the reason why these definitions are subscribed with A and the spaces of densities are also denoted with A's.

Definition 3.1 *Let $f, g \in \mathcal{A}_\ell^2$ be any two probability density functions on $(-\ell, \ell)$. We define its inner product as the function $<,>_A: \mathcal{A}_\ell^2 \times \mathcal{A}_\ell^2 \to \mathbb{R}$ given by*

$$< f, g >_A = \frac{1}{4\ell} \int_{-\ell}^{\ell} \int_{-\ell}^{\ell} \log \frac{f(x)}{f(y)} \times \log \frac{g(x)}{g(y)} \, dxdy \,. \tag{3}$$

Developing the right hand side of (3) we get

$$\begin{aligned} < f, g >_A &= \frac{1}{4\ell} \int_{-\ell}^{\ell} \int_{-\ell}^{\ell} \Big[\log f(x) - \log f(y)\Big] \times \Big[\log g(x) - \log g(y)\Big] \, dxdy \\ &= \frac{1}{2\ell} \left[\int_{-\ell}^{\ell} \int_{-\ell}^{\ell} \log f(x) \log g(x) \, dxdy - \int_{-\ell}^{\ell} \int_{-\ell}^{\ell} \log f(x) \log g(y) \, dxdy \right] \\ &= \int_{-\ell}^{\ell} \log f(x) \log g(x) \, dx - \frac{1}{2\ell} \int_{-\ell}^{\ell} \log f(x) \, dx \int_{-\ell}^{\ell} \log g(y) \, dy. \end{aligned} \tag{4}$$

We point out that, for density functions in $\mathcal{A}_\ell^2$ satisfying $\int_{-\ell}^{\ell} \log f(x) dx = 0$, their inner product in $\mathcal{A}_\ell^2$ is equal to the ordinary L^2 inner product of their logarithms.

Definition 3.2 *Let $f \in \mathcal{A}_\ell^2$ be any probability density function on $(-\ell, \ell)$. We define its norm, associated with the inner product defined in (3), as the function $\| \, \|_A: \mathcal{A}_\ell^2 \to \mathbb{R}$ given by*

$$\begin{aligned} \| f \|_A = \sqrt{< f, f >_A} &= \left[\frac{1}{2\ell} \int_{-\ell}^{\ell} \int_{-\ell}^{\ell} \log^2 \frac{f(x)}{f(y)} \, dxdy \right]^{1/2} \\ &= \left[\int_{-\ell}^{\ell} \log^2 f(x) \, dx - \frac{1}{2\ell} \left(\int_{-\ell}^{\ell} \log f(x) \, dx \right)^2 \right]^{1/2} . \end{aligned} \tag{5}$$

Definition 3.3 *Let $f, g \in \mathcal{A}_\ell^2$ be any two probability density functions on $(-\ell, \ell)$. We define the distance between them as the function $d_A : \mathcal{A}_\ell^2 \times \mathcal{A}_\ell^2 \to \mathbb{R}$ given by*

$$d_A(f, g) = \left[\frac{1}{4\ell} \int_{-\ell}^{\ell} \int_{-\ell}^{\ell} \left(\log \frac{f(x)}{f(y)} - \log \frac{g(x)}{g(y)} \right)^2 \, dxdy \right]^{1/2} . \tag{6}$$

The existence of the inner product in $\mathcal{A}_\ell^2$ is guaranteed by the boundness of the functions and their logarithms. From (3), (5) and (6) next theorem immediately follows

Theorem 3.1 *The set of probability density functions with the inner product, norm and distance defined previously, $(\mathcal{A}_\ell^2, <,>_A, || \, ||_A, d_A)$, is a pre-Hilbert space (also metric and normed space).*

In order to complete $\mathcal{A}_\ell^2$ to a Hilbert space of probability densities we need to characterize the closure of $\mathcal{A}_\ell^2$, i.e. to identify all the limits of Cauchy's sequences of elements in $\mathcal{A}_\ell^2$. A comfortable way to do it is to build up a Hilbert base in $\mathcal{A}_\ell^2$ and then obtain the Fourier coefficients of the probability densities. The completion of $\mathcal{A}_\ell^2$ is then got when all absolutely square-sumable sequences of Fourier coefficients are considered (Berberian, 1961, p. 49-50).

Now, we state and prove the following theorem.

Theorem 3.2 *If $\{\varphi_j\}_{j\geq 0}$ is a set of bounded functions that are orthonormal and a Hilbert base of $L^2(-\ell,\ell)$ with $\varphi_0(x)=\dfrac{1}{\sqrt{2\ell}}$, then $\{\psi_j\}_{j\geq 1}$ where $\psi_j=\mathcal{C}\left[\exp(\varphi_j)\right],$ is an orthonormal set in $\mathcal{A}_\ell^2$.*

Proof. We claim that for all $j\geq 1$, is $\psi_j\in\mathcal{A}_\ell^2$. In fact, since φ_j is bounded, then $0<m<\psi_j<M$. That is, ψ_j is also bounded. Furthermore, $\displaystyle\int_{-\ell}^{\ell}\exp[\varphi_j(x)]\,dx<+\infty$, and $\displaystyle\int_{-\ell}^{\ell}\psi_j(x)\,dx=\int_{-\ell}^{\ell}\mathcal{C}\left[\exp(\varphi_j(x)\right]\,dx=1$. As a consequence, $\parallel\psi_j\parallel_A^2<+\infty$, and our claim is proved.

Next, orthogonality is established, i.e., $<\psi_j,\psi_k>_A=0$ for all (j,k) with $j\neq k$. Explicitly,

$$
\begin{aligned}
<\psi_j,\psi_k>_A &= \int_{-\ell}^{\ell}\log\left\{\mathcal{C}(\exp[\varphi_j(x)])\right\}\log\left\{\mathcal{C}(\exp[\varphi_k(x)])\right\}\,dx\\
&-\frac{1}{2\ell}\int_{-\ell}^{\ell}\log\left\{\mathcal{C}(\exp[\varphi_j(x)])\right\}\,dx\int_{-\ell}^{\ell}\log\left\{\mathcal{C}(\exp[\varphi_k(y)])\right\}\,dy\\
&=\int_{-\ell}^{\ell}\varphi_j(x)\varphi_k(x)\,dx-\int_{-\ell}^{\ell}\varphi_j(x)\log\left[\int_{-\ell}^{\ell}\exp[\varphi_k(w)]\,dw\right]\,dx\\
&-\int_{-\ell}^{\ell}\varphi_k(x)\log\left[\int_{-\ell}^{\ell}\exp[\varphi_j(w)]\,dw\right]\,dx\\
&+2\ell\log\left[\int_{-\ell}^{\ell}\exp[\varphi_k(w)]\,dw\right]\log\left[\int_{-\ell}^{\ell}\exp[\varphi_j(z)]\,dz\right]\\
&-\frac{1}{2\ell}\int_{-\ell}^{\ell}\log\left\{\frac{\exp[\varphi_j(x)]}{\displaystyle\int_{-\ell}^{\ell}\exp[\varphi_j(z)]\,dz}\right\}\,dx\int_{-\ell}^{\ell}\log\left\{\frac{\exp[\varphi_k(x)]}{\displaystyle\int_{-\ell}^{\ell}\exp[\varphi_k(w)]\,dw}\right\}\,dx\\
&=-\frac{1}{2\ell}\int_{-\ell}^{\ell}\varphi_j(x)\,dx\int_{-\ell}^{\ell}\varphi_k(x)\,dx=0.
\end{aligned}
$$

Notice that in the last equality we have used the fact that for all $j\geq 1$, is $<\varphi_j,\varphi_0>=0$, the orthogonality in $L^2(-\ell,\ell)$. That is, $\displaystyle\int_{-\ell}^{\ell}\varphi_j(x)\,dx=0$. This completes the proof. □

Following the notation in the preceding theorem,

Theorem 3.3 *If $g\in\mathcal{A}_\ell^2$, then $\displaystyle\sum_{j=1}^{\infty}\left|<g,\psi_j>_A\right|^2<+\infty$.*

Proof. In fact,

$$
\begin{aligned}
\sum_{j=1}^{\infty}\Big|< g,\psi_j >_A\Big|^2 &= \sum_{j=1}^{\infty}\left|< \log g, \log\psi_j >_A - \frac{1}{2\ell}\int_{-\ell}^{\ell}\log g(x)\,dx\int_{-\ell}^{\ell}\log\psi_j(y)\,dy\right|^2 \\
&= \sum_{j=1}^{\infty}\Big|< \log g, \varphi_j - \log\int_{-\ell}^{\ell}\exp[\varphi_j(z)]\,dz > \\
&- \frac{1}{2\ell}\left(\int_{-\ell}^{\ell}\log g\,dx\right)\left(\int_{-\ell}^{\ell}\varphi_j(x)\,dx - 2\ell\ \log\int_{-\ell}^{\ell}\exp([\varphi_j(z)]\,dz\right)\Big|^2 \\
&= \sum_{j=1}^{\infty}\Big|< \log g, \varphi_j > - \left(\int_{-\ell}^{\ell}\log g\,dx\right)\log\int_{-\ell}^{\ell}\exp[\varphi_j(z)]\,dz \\
&- \frac{1}{2\ell}\left(\int_{-\ell}^{\ell}\log g\,dx\right)\left(\int_{-\ell}^{\ell}\varphi_j(x)\,dx - 2\ell\ \log\int_{-\ell}^{\ell}\exp[\varphi_j(z)]\,dz\right)\Big|^2 \\
&= \sum_{j=1}^{\infty}\Big|< \log g, \varphi_j >\Big|^2 < +\infty.
\end{aligned}
$$

□

After obtaining an orthogonal set in $\mathcal{A}_\ell^2$, we can complete $\mathcal{A}_\ell^2$ automatically by closing the space with all density functions whose Fourier coefficients are square-summable. This closed space is denoted $A^2(-\ell,\ell)$ and it is a Hilbert space.

Perturbation (1) is extended from $\mathcal{A}_\ell^2$ to $A^2(-\ell,\ell)$ in a standard way. Any two densities $f, g \in A^2(-\ell,\ell)$ are characterized by Cauchy sequences of bounded densities in $\mathcal{A}_\ell^2$, namely $\{f_i\}_{i\geq 1}$ and $\{g_i\}_{i\geq 1}$. Perturbation is defined for i-th term, $f_i \oplus g_i = h_i$. The sequence $\{h_i\}_{i\geq 1}$ is easily proved to be a Cauchy sequence in $\mathcal{A}_\ell^2$ and then representing a density $h \in A^2(-\ell,\ell)$. The extension is thus attained by setting $f \oplus g = h$. Extensions of power transformation (2), inner product (3), norm (5) and distance (6) are defined in a similar way.

However, the density functions of $A^2(-\ell,\ell)$ are not explicitly characterized. The following theorem gives an equivalent explicit definition of $A^2(-\ell,\ell)$.

Theorem 3.4 *Let $g : (-\ell,\ell) \to \mathbb{R}$ be a non-negative function such that $\int_{-\ell}^{\ell} g(x)\,dx = 1$. Then, $g \in A^2(-\ell,\ell)$ if and only if $\log g \in L^2(-\ell,\ell)$.*

Proof. The result follows from $\left|\int_{-\ell}^{\ell}\log g(x)\,dx\right|^2 \leq \int_{-\ell}^{\ell}\left|\log g(x)\right|^2 dx$. □

A simple density function in $(-1/2, 1/2)$ is the Beta density, namely

$$f(x) = \frac{1}{B(a,b)}\left(x+\frac{1}{2}\right)^{a-1}\left(x-\frac{1}{2}\right)^{b-1},\ a>0,\ b>0,$$

where $B(a,b)$ is the Euler Beta function. It has either zeros or asymptotes at $\pm 1/2$ and then it is not in $\mathcal{A}_{1/2}^2$. However, $\log f(x)$ is in $L^2(-1/2,1/2)$ and then $f \in A^2(-1/2,1/2)$.

4 Hilbert Basis in $A^2(-\ell,\ell)$

Theorem 3.2 provides a simple method to find out Hilbert basis in $A^2(-\ell,\ell)$. Each bounded Hilbert base in $L^2(-\ell,\ell)$ containing the constant is straightforward transformed into Hilbert base

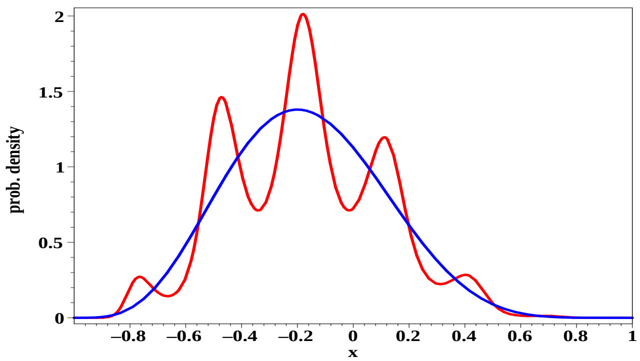

Figure 1. Blue: Beta density in $(-1,1)$, $a=5$, $b=7$. Red: 12 term truncated Fourier series with respect the Fourier base in $A^2(-1,1)$.

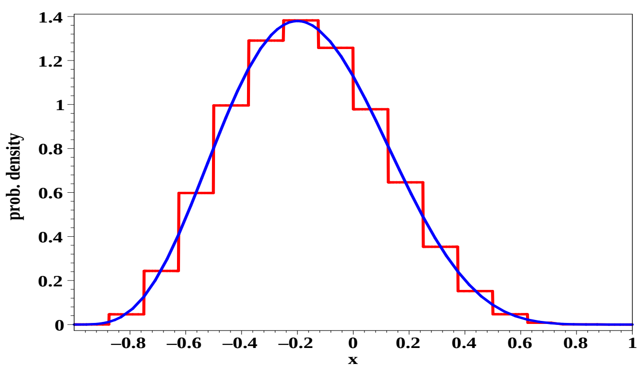

Figure 2. Blue: Beta density in $(-1,1)$, $a=5$, $b=7$. Red: 15 term truncated Fourier series with respect the Haar base in $A^2(-1,1)$.

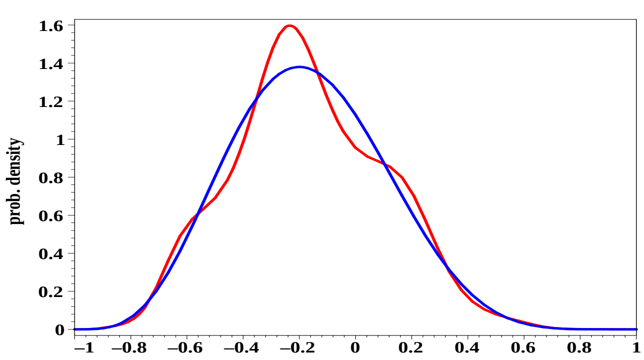

Figure 3. Blue: Beta density in $(-1,1)$, $a=5$, $b=7$. Red: 12 term truncated Fourier series with respect the Legendre base in $A^2(-1,1)$.

in $A^2(-\ell,\ell)$ taking closed exponentials. Three standard Hilbert basis for $A^2(-\ell,\ell)$ follow:

Fourier base

For $i=1,2$ and $k\geq 1$, $\{f_{in}(x)\}$ where

$$f_{1k}(x)=\mathcal{C}\left[\exp\left(\frac{1}{\sqrt{\ell}}\cos\frac{k\pi x}{\ell}\right)\right]\ ,\ f_{2k}(x)=\mathcal{C}\left[\exp\left(\frac{1}{\sqrt{\ell}}\sin\frac{k\pi x}{\ell}\right)\right]\ ,$$

is a Fourier basis in $A^2(-\ell,\ell)$.

Haar base

For $m=0,1,2,\dots$ and $n=0,1,2,\dots,2^m-1$,

$$\psi_{mn}(x)=\mathcal{C}\left[\exp\left\{\sqrt{\frac{2^m}{2\ell}}\ \psi\left(2^m\frac{x+\ell}{2\ell}-n\right)\right\}\right]\ ,$$

where

$$\psi(x)=\begin{cases}1, & 0\leq x<\frac{1}{2};\\ -1, & \frac{1}{2}\leq x<1;\\ 0, & \text{otherwise,}\end{cases}$$

is the Haar (1910) function. It is a Haar base in $A^2(-\ell,\ell)$.

Legendre base

For $n\geq 1$,

$$\xi_n(x)=\mathcal{C}\left[\exp\left[\sqrt{\frac{2n+1}{2\ell}}P_n\left(\frac{x}{\ell}\right)\right]\right]$$

where P_n are the ordinary Legendre polynomials (Abramowitz, 1972, p. 775), i.e., $P_0(z)=1$, $P_1(z)=z$, $P_2(z)=(3z^2-1)/2$, and

$$P_n(z)=\frac{1}{2^n}\sum_{k=0}^{\lfloor n/2\rfloor}(-1)^k\binom{n}{k}\binom{2n-2k}{n}z^{n-2k},\ n=1,2,3,\dots$$

is a Legendre basis in $A^2(-\ell,\ell)$.

A Hilbert base in $A^2(-\ell,\ell)$ allows us representation of densities by Fourier series. Assume $\{\psi_i\}_{i\geq 1}$ is a Hilbert base in $A^2(-\ell,\ell)$ and $f\in A^2(-\ell,\ell)$. The i-th Fourier coefficient of f is $c_i=\langle f,\psi_i\rangle_A$ and f is represented by the Fourier series

$$f(x)=\bigoplus_{i=1}^{\infty}(c_i\otimes\psi_i)\ .$$

Truncated Fourier series produce least squares approximations of probability densities in the sense of Aitchison square norm, i.e. the error is Aitchison-orthogonal to the projection subspace. Figures 1, 2 and 3 show a beta probability density in $(-1,1)$ with parameters $a=5$, $b=7$ and a truncated Fourier series using the Fourier (Fig. 1), Haar (Fig. 2) and Legendre (Fig. 3) Hilbert basis. We notice that Aitchison norm approximation forces truncated series to match more accurately low values of the density than higher ones as we expect when approximating compositions in the simplex following Aitchison geometry.

5 Conclusion

Basic concepts of Aitchison geometry of the simplex have been generalized to the set of probability density functions with support on a finite interval. Probability density functions are intuitively viewed as compositions of infinitely many parts and the operations in the simplex are in this way adapted to density functions.

The set of probability density functions whose logarithm is square-summable is shown to be an infinite dimensional Hilbert space and some Hilbert basis are built up from the corresponding well known Hilbert basis in the space of the square-summable functions.

The obtained results suggest several research lines concerning probability density functions. In this context, appealing points are the possibility of representing probability densities by their Fourier coefficients and the review of the concept of expectation of a random probability density (predictives in Bayesian statistics).

We have avoided the problem of generalizing this theory to densities whose support is the whole real line. This generalization implies additional difficulties and it remains open.

6 References

Abramowitz, M. and Stegun, I.A., 1972, Handbook of Mathematical Functions: Dover, New York.

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

Aitchison, J., Barceló-Vidal, C., Egozcue, J.J. and V. Pawlowsky-Glahn, 2002, A concise guide to the algebraic-geometric structure of the simplex, the sample space for compositional data analysis. Proceedings of IAMG'02 — The seventh annual conference of the International Association for Mathematical Geology, September 2002, Berlin.

Berberian, S.K., 1961, Introduction to Hilbert Space: University Press, New York.

Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G. and Barceló-Vidal, C., 2003, Isometric logratio transformations for compositional data analysis: Mathematical Geology, v. 35, no. 3, p. 279-300.

Haar, A., 1910, Zur Theorie der Ortogonalen Funktionen-Systeme: Math. Ann., v. 69, p. 331-371.