Simplicial splines for representation of density functions

J. Machalová[1], K. Hron[1], R. Talská[1], A. Menafoglio[2]
[1] Palacký University, Olomouc, Czech Republic; *jitka.machalova@upol.cz*
[2] Politecnico di Milano, Milano, Italy

Abstract

Probability density functions are popularly known as non-negative functions with unit integral constraint. This clearly inhibit their direct processing using standard methods of functional data analysis (Ramsay and Silverman, 2005) since unconstrained functions are assumed there. In addition, density functions are rather characterized by deeper geometrical features that need to be taken into account for any reliable analysis (Egozcue et al., 2006; van den Boogaart, 2014). If we restrict to bounded support $I = [a, b] \subset \mathbf{R}$ that is mostly used in practical applications, density functions can be represented with respect to Lebesque reference measure using the Bayes space $\mathcal{B}^2(I)$ of functions with square-integrable logarithm. The Bayes space $\mathcal{B}^2(I)$ has structure of separable Hilbert space that enables to construct an isometric isomorphism between $\mathcal{B}^2(I)$ and $L^2(I)$, the L^2 space restricted to I. An isometric isomorphism between $\mathcal{B}^2(I)$ and $L^2(I)$ is represented by the *centred log-ratio* (clr) transformation (van den Boogaart et al., 2014). The clr transformation induces an additional zero-integral constraint that needs to be taken into account for computation and analysis on clr-transformed density functions. As the clr space is clearly a subspace of $L^2(I)$, hereafter it is denoted as $L^2_0(I)$.

Functional data analysis relies strongly on approximation of the input functions using splines (Ramsay and Silverman, 2005). However, splines are mostly utilized just from this perspective, without considering further methodological consequences. Because statistical processing of density functions requires a deeper geometrical background, provided by the Bayes spaces, this should be followed also by the respective spline representation, performed preferably in the clr space $L^2_0(I)$. In Machalová et al. (2016) and Talská et al. (2017) a first attempt of constructing a spline representation that would honor the zero integral constraint of clr transformed densities was performed. The problem is that B-splines that form basis for the spline expansion come from $L^2(I)$, but not from $L^2_0(I)$. The contribution presents an important step ahead - such B-splines are constructed that are functions in the clr space $L^2_0(I)$. Consequently, the B-splines can be expressed also directly in $\mathcal{B}^2(I)$ and the spline representation written in terms of the Bayes space; hereafter we refer to *simplicial splines*. Apart from purely methodological advantages, using simplicial splines simplifies construction and interpretation of spline coefficients that can be considered as coefficients of a (orthonormal) basis in $\mathcal{B}^2(I)$. Projecting of a density function on subset of such B-splines provides an opportunity to a range of practical applications.

References

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

Egozcue, J.J., Díaz-Barrero, J.L., Pawlowsky-Glahn, V. (2006). Hilbert space of probability density functions based on Aitchison geometry. *Acta Mathematica Sinica (English Series) 22*(4), pp. 1175–1182.

Machalová, J., Hron, K., Monti, G. S. (2016). Preprocessing of centred logratio transformed density functions using smoothing splines. *Journal of Applied Statistics 43*(8), pp. 1419–1435.

Ramsay, J. Silverman, B.W. (2005). *Functional Data Analysis.* Heidelberg: Springer.

Talská R., Menafoglio, A., Machalová, J., Hron, K., Fišerová, E. (2017). Compositional regression with functional response. *In preparation.*