ALGEBRAIC METHOD ON EXPERIMENTAL DESIGN INI CAMBRIDGE UK OCT 25-26 2011

HILBERT BASIS IN DESIGN AND LOGLINEAR MODELS

GIOVANNI PISTONE

1. Hilbert bases of orthogonal designs

Consider three binary variables X_1, X_2, X_3 with levels $+1, -1$. The table of interactions is

	000	001	010	011	100	101	110	111
$+++$	1	1	1	1	1	1	1	1
$++-$	1	−1	1	−1	1	−1	1	−1
$+-+$	1	1	−1	−1	1	1	−1	−1
$+--$	1	−1	−1	1	1	−1	−1	1
$-++$	1	1	1	1	−1	−1	−1	−1
$-+-$	1	−1	1	−1	−1	1	−1	1
$--+$	1	1	−1	−1	−1	−1	1	1
$---$	1	−1	−1	1	−1	1	1	−1

The matrix K excluding the constant and the 3-way interaction is

(1) $K =$

	001	010	011	100	101	110
$+++$	1	1	1	1	1	1
$++-$	−1	1	−1	1	−1	1
$+-+$	1	−1	−1	1	1	−1
$+--$	−1	−1	1	1	−1	−1
$-++$	1	1	1	−1	−1	−1
$-+-$	−1	1	−1	−1	1	−1
$--+$	1	−1	−1	−1	−1	1
$---$	−1	−1	1	−1	1	1

Consider the lattice

$$OA_{3,2} = \ker(K^t) \cap \mathbb{Z}_{\geq}^{2^3}$$

Each element of $OA_{3,2}$ is the counting function of an orthogonal array.

The lattice has a basis whose elements are the rows of the matrix

$$H = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

i.e. the two fractions 2^{3-1} which are generated by $x_1x_2x_3 = \pm 1$.

This has been done up to five factors in [2] to provide a full classification of orthogonal designs. Computations with six or more factors were unfeasible on `4ti2` [1].

2. Hilbert bases in lattice exponential families

Consider the exponential family on $\mathcal{X} = \{+1, -1\}^3$ with unnormalized density

$$q(x;\theta) = \exp(\theta x_1 x_2 x_3) \tag{2}$$

The model matrix is is given by the 3-way interaction

$$A = [1 - 1 - 1 + 1 - 1 + 1 + 1 - 1]^t$$

The matrix K is the basis of the lattice $\operatorname{span}(A) \cap \mathbb{Z}_{\geq}^{2^3} = OA_{3,2}$. Note that K is boolean.

It follows from [3], [7], [6] that the matrix K describes the limits of the exponential family (2) as $\theta \to \pm\infty$.

3. Computer experiments

Viceversa, the statement about the border of an exponential family could be re-phrased by saying the the exponential family (2) is a *relaxation* or continuous interpolation of the two fractional designs 2^{3-1}.

The unnormalized densities of the eaponential family (2) satisfy all the binomial equations derived from from the matrix K in (1). Moreover, at least on this example,

$$\mathbb{E}_\theta(X_i) = 0, \quad 1 = 1, 2, 3.$$

The (continuous) sampling of the exponential family has propeties similar to the sampling on the orthogonal fractions.

This suggest the use of the relazation form of the orthomonal fractions in computer experiments:

(1) The sampling with fractional factorial is considered inappropriate in computer experiments and could be replaced by Gibbs sampling of the relaxation family.
(2) Given any funtion $f : \mathcal{X} \to \mathbb{R}$ any algorithm of gradient desent on the funtion $\theta \mapsto \mathbb{E}_\theta(f)$ will approach the border of the relaxation family and eventually sample from an orthogonal array with high probability. Such algorithms are discussed e.g. in [4], [5]

This is a joint project with Henry Wynn.

References

1. 4ti2 team, *4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces*, Available at www.4ti2.de.
2. Enrico Carlini and Giovanni Pistone, *Hibert bases for orthogonal arrays*, Journal of Statistical Theory and Practice (2007), 299–309.
3. Thomas Kahle, *On Boundaries of Statistical Models*, Ph.D. thesis, Fakultät der Mathematik und Informatic, Universität Leipzig, 2010.
4. Luigi Malagò, Matteo Matteucci, and Bernardo Dal Seno, *An information geometry perspective on estimation of distribution algorithms: boundary analysis*, GECCO '08: Proceedings of the 2008 GECCO conference companion on Genetic and evolutionary computation (New York, NY, USA), ACM, 2008, pp. 2081–2088.
5. Luigi Malagò, Matteo Matteucci, and Giovanni Pistone, *Towards the geometry of estimation of distribution algorithms based on the exponential family*, Foundations of Genetic Algorithms XI. Schwarzenberg, Austria, 59 January 2011 (H-G Beyer and W. B. Langdon, eds.), 2010, Pre-proceedings, pp. 221–232.
6. Luigi Malagò and Giovanni Pistone, *A note on the border of an exponential family*, arXiv:1012.0637v1, 2010.
7. J. Rauh, T. Kahle, and N. Ay, *Support sets in exponential families and oriented matroid theory*, Proceedings WUPES'09, invited for special issue of IJAR arXiv:0906.5462, 2009.

Collegio Carlo Alberto, Moncalieri Italy