Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 138 (2008) 259–271
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Algebraic solutions to the connectivity problem for m-way layouts: Interaction-contrast aliasing

H.P. Wynn*

Department of Statistics, London School of Economics, London, UK

Available online 13 May 2007

Abstract

An m-way layout, or design, is called complete if there is at least one observation per cell. Without any side conditions the parameters are *not* estimable but contrasts may be estimable and they are all estimable for the complete layout. For $m = 2$ the condition for estimability is the well-known connectivity condition: one can "walk" from any row to any other row, stepping on columns. The case $m > 2$ remains unsolved in some sense. The principal method used here is the Gröbner basis (G-basis) method introduced by Pistone and Wynn [1996. Generalised confounding with gröber bases. Biometrica 83, 653–666]. The problem is set up, using indicator functions, and necessary and sufficient conditions given for full estimability and various constructions using the G-basis method.

Keywords: Experimental design; m-way layout; Connectivity; Aliasing; Gröbner basis

1. The connectivity problem

The m-way layout, or table, is of considerable importance in statistics. A layout is complete if every cell has at least one observation. But incomplete tables are more complex entities. There are classical experimental designs, such as balanced incomplete blocks, designed to be incomplete, but for very incomplete tables there may be a lack of identifiability. In the language of linear models some *contrasts* may not be estimable. The condition for estimability for the 2-way table ($m=2$) is well known, namely the connectivity of the table; we shall discuss the details shortly. However, for $m > 2$ the conditions are more complex. This paper is a contribution to understanding this more complex problem. One might describe the problem as unsolved, but "not fully understood" would, perhaps, be a better description. We try, here, to give some insight taking advantage of the Gröbner basis (G-basis) method introduced by Pistone and Wynn (1996) and developed in Pistone et al. (2001).

Considerable effort has been devoted to the case of three factors: rows, columns and treatments. The interesting constructive methods of J.D and E.J Godolophin (Godolphin, 2004; Godolphin and Godolphin, 2001) should be mentioned. Earlier work is well represented by the important paper of Dodge et al. (1976). Ghosh (1989) developed a graphic method with a similar algebraic background. Weeks and Williams (1964) paper was one of the first papers

* Tel.: +44 020 7955 6063; fax: +44 020 7955 7416.
E-mail address: h.wynn@lse.ac.uk

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2007.05.007

to set the basic problem up and giving appropriate rank-based solutions. We can find one monograph: Butz (1982). Park and Shah (1995) and John and Williams (1995) cover the important block design case. The author has made some previous contributions: Wynn (1977) gives a solution in a special case and Dodge and Wynn (1988) cover the case $m=3$ with connections to the last section of this paper.

The classical way to set up an m-way additive models is as follows. Let there be m independent factors $j=1,\dots,m$ where factor j has levels $i_j=1,\dots,n_j$. The factors form an m-way table with cells labelled $(i_1,\dots,i_m)$. We consider the single replicate case where there is at most one observation per cell. We here call a cell "live" if it holds an observation $Y_{i_1,\dots,i_m}$. Thus, a layout (table, design) is defined to be complete if and only if all cells are live.

For live cells the classical model is

$$Y_{i_1,\dots,i_m}=\mu+\theta^{(1)}_{i_1}+\theta^{(2)}_{i_2}+\cdots+\theta^{(m)}_{i_m}+\varepsilon_{i_1,\dots,i_m}, \tag{1}$$

where the errors, $\varepsilon_{i_1,\dots,i_m}$, are uncorrelated with zero mean and non-zero, finite constant variance. Note that the estimability conditions of this paper are not dependent on constancy of variance.

A linear function of the parameters is defined to be estimable if it possesses a linear unbiased estimator. It is well known that in general, without further side assumptions, for a factor j the individual parameters $\theta^{(j)}_{i_j}$ are not estimable but that linear contrasts among the parameters, defined to be of the form

$$\sum_{i=1}^{n_j} c_i\theta^{(j)}_i, \tag{2}$$

with $\sum_i c_i=0$, may or may not be estimable. This drives our definition.

Definition 1. In an m-way layout an index set $J\subseteq\{1,\dots,m\}$ with standard additive model (1) layout is called J-connected if all the individual factor (2) contrasts for factors $j\in J$ are estimable. If all factor contrasts are estimable, that is $J=\{1,\dots,m\}$, we call the design totally connected.

Note that in defining the term J-connected we are *not* restricting the full underlying model to just the factors in J. Rather, we retain the full model so that we are considering the factors in J, *in the presence of* the factors in the complement $J^{\mathrm{c}}=\{1,\dots,m\}\backslash J$. It is possible that a design can be J-connected without being totally connected. Note also that for a factor j to be estimable it is enough that all elementary contrasts of the form $\theta^{(j)}_i-\theta^{(j)}_k$, $i\neq k$ are estimable.

The condition for total connectivity in the case $m=2$ is well known (see e.g. Weeks and Williams, 1964) and motivates Definition 1.

Theorem 2. *A 2-way layout $(m=2)$ is totally connected if every cell is connected to every other cell with a path which makes "steps" on live cells, moving in the rows (factor 1) or columns (factor 2) separately.*

The follow 3×3 table $(m=2, n_1=n_2=3)$ is connected. The blank cells correspond to non-live cells and we have placed the observation in the live cells:

Y_{11}	Y_{12}	
	Y_{22}	Y_{23}
		Y_{33}

As an example, an unbiased linear estimator of the row contrast $\theta^{(1)}_1-\theta^{(1)}_3$ is $Y_{12}-Y_{22}+Y_{23}-Y_{33}$.

In this paper, we adopt a somewhat algebraic approach, the starting point for which is the recasting of the model in terms of indicator functions. Thus, define the functions

$$x^{(j)}_i,\quad i=1,\dots,n_j,\quad j=1,\dots,m$$