

Indicator functions and saturated fractions for factorial designs: a case study

Funzioni indicatrici e frazioni sature per piani fattoriali: un case study

Roberto Fontana, Fabio Rapallo and Maria Piera Rogantin

Abstract We consider a recently developed algebraic criterion to check whether a fraction is saturated or not for a given model. Such criterion is based on combinatorial algebraic objects, namely the circuit basis of the design matrix of the model. We show on a case study how to use indicator functions of the circuits to classify saturated fractions. The connection with two optimality criteria is also analyzed.

Abstract Recentemente è stato messo a punto un criterio algebrico per verificare se una frazione è satura rispetto ad un modello assegnato. Tale criterio è basato sui circuiti associati alla matrice del piano fattoriale. Viene illustrato mediante un case study l'uso delle funzioni indicatrici dei circuiti per classificare le frazioni sature. Viene anche analizzato il collegamento con due criteri di ottimalità.

Key words: experimental design, circuits, optimality, Generalized Word Length Pattern

1 Introduction

Saturated fractions have the minimum number of points to estimate all the parameters of a given model. Consequently there are no degrees of freedom to estimate the error term. They are of common use in sciences and engineering for highly ex-

Roberto Fontana

Department DISMA, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10127 TORINO, Italy, e-mail: roberto.fontana@polito.it

Fabio Rapallo

Department DISIT, Università del Piemonte Orientale, Viale Teresa Michel 11, 15121 ALESSANDRIA, Italy, e-mail: fabio.rapallo@unipmn.it

Maria Piera Rogantin

Department DIMA, Università di Genova, Via Dodecaneso 35, 16146 GENOVA, Italy, e-mail: rogantin@dima.unige.it

pensive experiments. For general reference in design of experiments, the reader can refer to [8] and [1], where the issue of saturated fractions is discussed. In recent literature ([3], [5] and [4]) an algebraic and combinatorial description of saturated fractions of factorial designs has been presented. Given a model with p estimable parameters on the full factorial design, a new criterion is available in order to check whether a given fraction with p points is saturated or not. Such criterion is based on purely combinatorial properties of the fractions. Given the circuits associated to the design matrix, it checks whether the fraction contains some supports of the circuits or not.

We use here a methodology based on indicator functions ([2] and [6]) to highlight the symmetries and to classify both circuits and saturated fractions into few representative cases. Moreover, we show that such a combinatorial characterization gives some information about the D -optimality of the fractions, and we present some results also for fractions with more than p design points. In this work, we discuss these issues on a special example, namely the 2^4 design under the model with main factors and 2-way interactions. Indeed, this example is small enough for a complete enumeration and classification of all circuits and saturated fractions. Moreover, the structure of them reduces to few cases, due to the symmetry of the problem.

As described in [3], the key ingredient to characterize the saturated fractions of a factorial design is its circuit basis. Due to space limitations, we leave out the formal definition of circuit basis. The interested reader can refer to [7]. Given a model matrix $X_{\mathcal{D}}$ of a full factorial design $\mathcal{D}$, the circuit basis of $X_{\mathcal{D}}^t$ is a special basis of the kernel of $X_{\mathcal{D}}^t$ with major combinatorial properties, and is denoted with $\mathcal{C}$. It is known that $\mathcal{C}$ is always finite. For small- and medium-sized problems, the circuit basis $\mathcal{C}$ can be actually computed with symbolic softwares, such as 4ti2 [10].

We denote by $\text{supp}(f)$ the support of the vector f , i.e., the set of indices j such that $f_j \neq 0$.

Given a model matrix $X_{\mathcal{D}}$ on a full factorial design $\mathcal{D}$ with p estimable parameters, we recall that a fraction $\mathcal{F} \subset \mathcal{D}$ with p design points is saturated if $\det(X_{\mathcal{F}}) \neq 0$, where $X_{\mathcal{F}}$ is the restriction of $X_{\mathcal{D}}$ to the design points in $\mathcal{F}$. With a slight abuse of notation, $\mathcal{F}$ denotes both a fraction and its support. Under these assumptions, in [3] is proved that $\mathcal{F}$ is a saturated fraction if and only if it does not contain any of the supports $\{\text{supp}(f_1), \dots, \text{supp}(f_L)\}$ of the circuits of $X_{\mathcal{D}}^t$.

2 Case study. 2^4 with main effects and 2-way interactions

Let us consider the 2^4 design and the model with main factors and 2-way interactions. As mentioned above, this example is not trivial and easy to handle. Under a full-rank parametrization, the design matrix $X_{\mathcal{D}}$ of the full design has 16 rows and 11 columns. The points are its row labels while the model terms are its column labels. As the matrix $X_{\mathcal{D}}$ has rank 11, we search for fractions with 11 points.

A direct computation shows that there are $\binom{16}{11} = 4,368$ fractions with 11 points: among them 3,008 are saturated, and the remaining 1,360 are not.

The circuits of the 11×16 matrix $X_{\mathcal{D}}^t$, are needed to find the saturated fractions. Using 4t i 2, we have obtained 140 circuits; the cardinalities of their supports are 8 in 20 cases, 10 in 40 cases, 12 in 80 cases. For more details refer to [3].

In order to classify both circuits and saturated fractions into few representative cases using the indicator function of a two-level fractional factorial designs, we briefly recall some topics about it. We briefly describe the indicator function of a fractional factorial design for two-level designs. We consider d factors coded as -1 and 1 . The full factorial design $\mathcal{D}$ is $\{-1, 1\}^d$. We denote by $\mathcal{L}$ the set $\{0, 1\}^d$ and by X^α , $\alpha \in \mathcal{L}$, a main factor or an interaction. Given a single-replicate fraction $\mathcal{F} \subseteq \mathcal{D}$, the indicator function of $\mathcal{F}$ written in polynomial form is

$$F(x) = \sum_{\alpha \in \mathcal{L}} b_\alpha X^\alpha(x) = \sum_{\alpha_i \in \{0,1\}} b_\alpha X_1^{\alpha_1}(x) \cdots X_d^{\alpha_d}(x) = \sum_{\alpha_i \in \{0,1\}} b_\alpha x_1^{\alpha_1} \cdots x_d^{\alpha_d}$$

where $x = (x_1, \dots, x_d)$ is a point of $\mathcal{D}$ and X_i , $i = 1, \dots, d$, denotes a main factor. Given a fraction $\mathcal{F} \subseteq \mathcal{D}$, the coefficients can be calculated as $b_\alpha = \frac{1}{2^s} \sum_{x \in \mathcal{F}} X^\alpha(x)$.

Circuits

From now on, for ease in presentation, we refer to the support of a circuit simply as the circuit. Computing the indicator functions for all the 140 circuits, we find 8 inequivalent circuits. It is known that two designs are isomorphic if one can be obtained from the other by relabeling the factors, reordering the runs, and switching the levels of factors. We denote by e a sign, $e \in \{-1, +1\}$, that correspond to change the factor codes.

1. Support cardinality equal to 8. There are 20 circuits corresponding to two inequivalent circuits, up to isomorphisms:

- a. 8 circuits with indicator function: $\frac{8}{16} + \frac{8}{16} e_i X_i$, $i = 1, 2, 3, 4$
- b. 12 circuits with indicator function: $\frac{8}{16} + \frac{8}{16} e_i X_i X_j$, $i = 1, \dots, 3$ $j = i + 1, \dots, 4$

They are all regular fractions. We will refer to these two types of circuits as $8a$ and $8b$ respectively.

2. Support cardinality equal to 10. There are 40 circuits corresponding to two inequivalent circuits, for $i, j, h, k = 1, 2, 3, 4$ and $i \neq j \neq h \neq k$:

- a. 8 circuits (type $10a$) with indicator function:

$$\begin{aligned} \frac{10}{16} + \frac{2}{16} (X_i e_j X_j + X_i e_h X_h + X_i e_k X_k + e_j X_j e_h X_h + e_j X_j e_k X_k + e_h X_h e_k X_k) \\ - \frac{6}{16} X_i e_j X_j e_h X_h e_k X_k \end{aligned}$$

Notice that X_i has not coefficient. Each choice of signs produces the same equivalent indicator functions.

- b. 32 circuits (type $10b$) with indicator function:

$$\begin{aligned} & \frac{10}{16} + \frac{2}{16} (e_i X_i + e_j X_j + e_h X_h) \\ & + \frac{2}{16} (e_i X_i e_j X_j + e_i X_i e_h X_h + e_j X_j e_h X_h) - \frac{6}{16} e_i X_i e_j X_j e_h X_h \end{aligned}$$

3. Support cardinality equal to 12. There are 80 circuits corresponding to two inequivalent circuits, for $n \in \{1, 2, 3, 4\}$, $j, h, k \in \{1, 2, 3, 4\} \setminus \{n\}$:

- a. 16 circuits (type 12a) with indicator function:

$$\begin{aligned} & \frac{12}{16} + \frac{2}{16} (e_1 X_1 + e_2 X_2 + e_3 X_3 + e_4 X_4) + \frac{4}{16} e_1 X_1 e_2 X_2 e_3 X_3 e_4 X_4 \\ & - \frac{2}{16} (e_1 X_1 e_2 X_2 e_3 X_3 + e_1 X_1 e_2 X_2 e_4 X_4 + e_1 X_1 e_3 X_3 e_4 X_4 + e_2 X_2 e_3 X_3 e_4 X_4) \end{aligned}$$

- b. 64 circuits (type 12b) with indicator function:

$$\begin{aligned} & \frac{12}{16} + \frac{2}{16} X_n (e_n - e_j X_j - e_h X_h - e_k X_k) - \frac{2}{16} e_n X_n e_j X_j e_h X_h e_k X_k \\ & - \frac{2}{16} (e_n X_n (e_j X_j e_h X_h + e_j X_j e_k X_k + e_h X_h e_k X_k) - 2 e_j X_j e_h X_h e_k X_k) \end{aligned}$$

Notice that for each choice of sign e_1, e_2, e_3, e_4 there are a circuit of type 12a and four circuits of type 12b, one for each choice of n .

Saturated fractions

We classify the saturated fraction with respect to two criteria of optimality and to their intersection with the circuits of the model matrix.

In [4] is shown that all the saturated fractions can be classified according to the determinant of the information matrix and that, up to isomorphisms, there are 188 *inequivalent* saturated fractions: 1 fraction with determinant equal to $9 \cdot 2^{20}$; 20 fraction with determinant equal to $4 \cdot 2^{20}$; 167 fraction with determinant equal to $1 \cdot 2^{20}$.

We present here another criterium of optimality, based on the coefficients b_α of the indicator function, that encode information about the aliasing of the fraction: the Generalized Minimum Aberration (*GMA*) criterion, see for instance [9] for a detailed presentation. It is based on the Generalized Word Length Pattern (*GWLP*) i.e. the d -tuple $(\gamma_1, \dots, \gamma_d)$ where each element γ_j measures the degree of aliasing of the interactions of order j (if $j = 1$ the main factors are considered). More precisely:

$$\gamma_j = \sum_{\|\alpha\|=j, \alpha \in \mathcal{L}} \left(\frac{b_\alpha}{b_0} \right)^2, \quad j = 1, \dots, d$$

and $\|\alpha\| = \sum_{k=1}^d \alpha_k$ is the order of the interaction. The *GMA* criterion consists in the sequential minimization of the *GWLP*.

Now we consider the intersection between the inequivalent saturated fractions and the circuits of the model matrix; more precisely we consider two types of intersection. First, the number of points in the intersection $\mathcal{F} \cap \text{supp}(f_i)$, with $i = 1, \dots, 140$. The number of points in such intersections ranges between 4 and 11. Second, we consider the intersection of each fraction with circuits to which one point has been deleted. We denote by *k-incomplete circuit* a circuit to which k points have been deleted. It is worthwhile noting that each saturated fraction contains five *1-incomplete circuits* of different type. The previous two types of intersection fraction/circuits are shown in Table 1, together with the determinant of $X_{\mathcal{F}}^t X_{\mathcal{F}}$ and the *GWLP* of the fraction.

Table 1 Saturated fractions classified by: number of points in the intersection with the circuits; intersection with the *1-incomplete circuits*; determinant of the information matrix; *GWLP*

#($\mathcal{F} \cap \text{supp}(f_i)$)								type of circuits					det	$GWLP$	#
4	5	6	7	8	9	10	11								
0	15	25	25	40	30	0	5	12a	12b	12b	12b	12b	$9 \cdot 2^{20}$	(4,6,36,9)	1
1	14	18	37	38	21	10	1	8b	10b	10b	10b	12b	$4 \cdot 2^{20}$	(4,14,36,1)	12
								8a	10a	10a	10a	12a	$4 \cdot 2^{20}$	(12,6,12,25)	4
								8a	10b	10b	10b	12b	$4 \cdot 2^{20}$	(12,6,28,9)	4
2	11	21	33	40	26	6	1	8b	8b	10a	10b	12b	$1 \cdot 2^{20}$	(4,22,20,9)	12
								8a	8b	10b	10b	12b	$1 \cdot 2^{20}$	(12,14,28,1)	12
								8a	8a	10b	10b	12a	$1 \cdot 2^{20}$	(20,6,20,9)	6
3	9	20	34	42	26	6	0	8b	8b	8b	10b	10b	$1 \cdot 2^{20}$	(4,30,20,1)	12
								8a	8b	8b	10a	10b	$1 \cdot 2^{20}$	(12,22,12,9)	24
								8a	8a	8b	10b	10b	$1 \cdot 2^{20}$	(20,14,20,1)	24
5	5	20	30	50	30	0	0	8a	8a	8b	8b	8b	$1 \cdot 2^{20}$	(20,30,4,1)	12
4	8	17	34	48	25	4	0	8a	8b	8b	8b	10b	$1 \cdot 2^{20}$	(12,30,12,1)	36
								8a	8a	8b	8b	10a	$1 \cdot 2^{20}$	(20,22,4,9)	12
								8a	8a	8a	8b	10b	$1 \cdot 2^{20}$	(28,14,12,1)	12
4	11	9	37	56	18	4	1	8a	8a	8a	8a	12a	$1 \cdot 2^{20}$	(12,30,12,1)	4
								8a	8b	8b	8b	12b	$1 \cdot 2^{20}$	(36,6,4,9)	1

We observe that the *D* and *GMA* criteria coincide and are strictly related to the classification with respect to the circuits. The optimal fraction is the intersection of its five *1-incomplete circuits*, all deriving from circuits with 12 points in the support. Given a unique choice of e_1, e_2, e_3, e_4 for all the 5 circuits the product of their indicator functions gives the indicator function of the optimal fraction with *GWLP* = (4, 6, 36, 9):

$$\begin{aligned} & \frac{1}{16} ((11 + e_1 X_1 + e_2 X_2 + e_3 X_3 + e_4 X_4) + 3e_1 X_1 e_2 X_2 e_3 X_3 e_4 X_4 \\ & - (e_1 X_1 e_2 X_2 + e_1 X_1 e_3 X_3 + e_1 X_1 e_4 X_4 + e_2 X_2 e_3 X_3 + e_2 X_2 e_4 X_4 + e_3 X_3 e_4 X_4) + \\ & - 3(e_1 X_1 e_2 X_2 e_3 X_3 + e_1 X_1 e_2 X_2 e_4 X_4 + e_1 X_1 e_3 X_3 e_4 X_4 + e_2 X_2 e_3 X_3 e_4 X_4)) \end{aligned}$$

Fractions with more than 11 points

The intersections between a fraction and the circuits can be easily studied also for fractions with more than 11 points. Also in this case, a direct enumeration shows that the D -optimality is closely related the position of the design points compared with the circuits. Limits in space does not allow to deal with a complete presentation of all cases, therefore we report in Table 2 only the results for the estimable fractions with 13 points, and we consider only the D -optimality criterion. In this case we can observe that the D -optimality is reached with the fractions that contain the supports of 2 circuits with the largest support cardinality. Finally, one can show that the D -optimal fractions are optimal also under the $GWLP$ criterion.

Table 2 Estimable fractions with 13 points classified by: number of points in the intersection with the circuits; determinant of the information matrix

	# of intersection								det	#
	5	6	7	8	9	10	11	12		
1	9	20	16	48	32	12	2		$112 \cdot 2^{20}$	10
2	8	16	23	41	38	12	0		$64 \cdot 2^{20}$	15
3	7	14	24	38	48	6	0		$48 \cdot 2^{20}$	10

References

1. Bailey, R.A.: Design of comparative experiments. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge (2008)
2. Fontana, R., Pistone, G., Rogantin, M.P.: Classification of two-level factorial fractions. J. Stat. Plann. Inference **87**(1), 149–172 (2000)
3. Fontana, R., Rapallo, F., Rogantin, M.P.: A characterization of saturated designs for factorial experiments. J. Stat. Plann. Inference **147**, 204–211 (2014)
4. Fontana, R., Rapallo, F., Rogantin, M.P.: D -optimal saturated designs: a simulation study. arXiv preprint arXiv:1401.0844 (2014)
5. Fontana, R., Rapallo, F., Rogantin, M.P.: Two factor saturated designs: cycles, Gini index and state polytopes. J. Stat. Theory Pract. **8**(1), 66–82 (2014)
6. Kotsireas, I.S., Koukouvinos, C., Rogantin, M.P.: Inequivalent Hadamard matrices via indicator functions. Int. J. Applied Math. **16**(3), 355–363 (2004)
7. Ohsugi, H.: A dictionary of Gröbner bases of toric ideals. In: Harmony of Gröbner bases and the modern industrial society, pp. 253–281. World Sci. Publ., Hackensack, NJ (2012)
8. Raktue, B.L., Hedayat, A., Federer, W.T.: Factorial designs. John Wiley & Sons Inc., New York (1981). Wiley Series in Probability and Mathematical Statistics
9. Sun, F., Chen, J., Liu, M.Q.: Connections between uniformity and aberration in general multi-level factorials. Metrika **73**(3), 305–315 (2011)
10. 4ti2 team: 4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces. Available at www.4ti2.de (2008)