Robustness of Fractional Factorial Designs through Circuits

Original
Robustness of Fractional Factorial Designs through Circuits / Fontana, Roberto; Rapallo, Fabio. - ELETTRONICO. - (2021), pp. 615-620. ((Intervento presentato al convegno SIS 2021.

Availability:
This version is available at: 11583/2920212 since: 2021-09-01T15:53:48Z

Publisher:
Pearson

Published
DOI:

Robustness of Fractional Factorial Designs through Circuits

Piani Fattoriali Frazionari Robusti attraverso i Circuiti

Roberto Fontana and Fabio Rapallo

Abstract Given a model we define the robustness of an experimental design as a function of the number of estimable minimal sub-fractions of it. We show how the circuit basis of the design matrix can be used to see if a minimal fraction is estimable or not and we describe an algorithm for finding robust fractions.
Abstract *Dato un modello si definisce la robustezza di un piano sperimentale in funzione del numero dei suoi sottoinsiemi minimali per cui tale modello risulta stimabile. Si dimostra che è possibile determinare se una frazione minimale è stimabile usando i circuiti della design matrix e si descrive un algoritmo per la ricerca di frazioni robuste.*

Key words: algebraic statistics, design of experiments, robust fractions

1 Robustness of a Design

In Design of Experiments, the choice of a design from a candidate set of runs is probably the most relevant problem, with a number of open questions from the point of view of both theory and applications. When searching for an efficient experimental designs, we aim to select a design in order to produce the best estimates of the relevant parameters for a given sample size. There are a lot of criteria for the selection of a design, both in the area of model-based designs (e.g., D-optimality and related criteria), and model-free designs (e.g., orthogonal arrays, space filling designs). In this work we focus on the model-based setting and we limit the analysis

Roberto Fontana
Dipartimento di Scienze Matematiche, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, e-mail: roberto.fontana@polito.it

Fabio Rapallo
Dipartimento di Economia, Università di Genova, via Vivaldi 5, 16126 Genova, e-mail: fabio.rapallo@unige.it

to Fractional Factorial Designs. In particular, we consider the notion of robustness of a design. This property is particularly important when the design may be incomplete, for instance when for some reasons one does not have the measurement for all planned design points. Fractional Factorial Designs with removed runs are studied in, e.g., [1], [7], [11]. A combinatorial analysis of the problem is introduced in [2], but in a model-free context.

Following the works by Ghosh ([4] and [5]), we define here the robustness in terms of the estimability of a given model on the basis of incomplete designs.

Let $\mathscr{D}$ be a large discrete set in $\mathbb{R}^m$ from which a small set $\mathscr{F}$, usually referred to as design or fraction, is to be selected. One standard example is to consider as candidate set $\mathscr{D}$ the cartesian product of the level sets of the m factors. The set $\mathscr{D}$, when thought of as a design in its own right, is referred to as a full factorial.

We point out that in our theory the coding of the level set is irrelevant, so that for the level set of a factor with s levels we can use $\{0,\ldots,s-1\}$ or the complex coding or any other coding that is considered appropriate.

A linear model on the candidate set $\mathscr{D}$ is written as:

$$\mathbf{y} = X_{\mathscr{D}}\beta + \varepsilon \,,$$

where $X_{\mathscr{D}}$ is the full-design model matrix with dimensions $K \times p$, β is the p-dimensional vector of the parameters, $\mathbb{E}(\mathbf{y}) = X_{\mathscr{D}}\beta$. In this work we assume that the design matrix $X_{\mathscr{D}}$ is full rank to simplify the presentation.

In the model-based approach to experimental design, the quality of the chosen design $\mathscr{F}$ is expressed by some properties of $X_{\mathscr{F}}$. Here we focus our attention on the notion of robustness. Robustness measures how many minimal fractions (i.e., fractions with as many points as the number of parameters) are estimable. A minimal estimable fraction is named as a saturated fraction.

Definition 1. We define the *robustness* of a fraction $\mathscr{F}$ with design matrix $X_{\mathscr{F}}$ as

$$r(X_{\mathscr{F}}) = \frac{\#\text{ saturated } \ \mathscr{F}_p}{\#\mathscr{F}_p} = \frac{\#\text{ saturated } \ \mathscr{F}_p}{\binom{n}{p}} \tag{1}$$

where $\mathscr{F}_p$ denotes a fraction with p runs and $\#\{\cdot\}$ denotes the cardinality of the set $\{\cdot\}$.

In particular a design $\mathscr{F}$ is *robust* if its robustness is equal 1, $r(X_{\mathscr{F}}) = 1$.

2 Circuits of the Design Matrix and Robustness

Let $X = X_{\mathscr{F}}$ be a model matrix on $\mathscr{F}$, and assume that X has integer entries. To simplify the notation, we drop the subscript $\mathscr{F}$ if there is no ambiguity. The matrix X has dimensions $K \times p$. Moreover, in order to match the common notation in Statistics with the notation in Commutative Algebra, we consider the matrix $A = X^T$, the transpose of the model matrix.

We are interested to use a special basis of the kernel of A as a lattice in $\mathbb{Z}^K$. A vector $\mathbf{u} = (u(1), \ldots, u(K))$ belongs to $\ker(A)$ if $A\mathbf{u} = 0$, or equivalently if $\mathbf{u}$ is orthogonal to $A^T = X$. The *support* of a vector $\mathbf{u} \in \mathbb{Z}^K$ is the set of indices i $(i = 1, \ldots, K)$ such that $u(i) \neq 0$. We denote the support of $\mathbf{u}$ with $\mathrm{supp}(\mathbf{u})$.

The circuits of A are the integer vectors of $\ker(A)$ with relatively prime entries and with minimal support. This means that if $\mathbf{u}$ is a circuit, then there does not exist another circuit $\mathbf{v}$ with $\mathrm{supp}(\mathbf{v}) \subset \mathrm{supp}(\mathbf{u})$. We denote the set of all circuits of A with $\mathscr{C}(A)$ (or $\mathscr{C}(X)$ if we refer to a design matrix $X = A^T$). The set $\mathscr{C}(A)$ is called the circuit basis of the matrix A. Among the properties of the circuit basis, we make use of the following:

1. $\mathscr{C}(A)$ is a basis of $\ker(A)$ as vector space.
2. Every vector $\mathbf{v} \in \ker(A)$ can be written as a non-negative rational combination of $(K-p)$ circuits
$$\mathbf{v} = \sum c_j \mathbf{u}_j \qquad c_j \in \mathbb{Q}^+, \; \mathbf{u}_j \in \mathscr{C}(A)$$
and each circuit in the decomposition above is sign-compatible with $\mathbf{v}$.
3. The support of a circuit has cardinality at most $(p+1)$.

For the proofs and for a detailed introduction to circuits in the context of Commutative Algebra and Combinatorics, see [8].

The circuit basis of an integer matrix A can be computed through several packages for symbolic computation. The computations presented in the present paper are carried out with `4ti2`, see [9]. `4ti2` can be used as an independent executable program or as a package of the Computer Algebra System `Macaulay2`, see [6]. For small designs the computations are performed in a few seconds at most, and the circuit basis in the output can be easily analyzed. For instance, if we want to compute the circuit basis for the full factorial 2^4 design with main effects and first order interactions, it is enough to run `4ti2`, input the design matrix and the circuit basis with 140 elements is computed in less than 0.1 seconds on a standard PC.

From the minimal support property in the definition of the circuits we can use the circuits to see if a minimal fraction (i.e., a fraction with exactly p points) is saturated or not. Given a set $\mathscr{F}$ of p column-indices of A, the sub-matrix $A_{\mathscr{F}}$ is non-singular if and only if $\mathscr{F}$ does not contain any of the supports of the circuits $\mathbf{u} \in \mathscr{C}(A)$. This result is proved and applied to the analysis of Fractional Factorial Designs in [3].

Now, we extend the analysis to fractions with more than p points. The key property here is that the circuit basis is consistent under selection of sub-fractions. Consider two fractions $\mathscr{F}_1$ and $\mathscr{F}_2$ with k_1 and k_2 design points respectively, such that $\mathscr{F}_1 \subset \mathscr{F}_2$. Without loss of generality, the matrix $X_{\mathscr{F}_2}$ can be partitioned into

$$X_{\mathscr{F}_2} = \begin{pmatrix} X_{\mathscr{F}_1} \\ X_{\mathscr{F}_2 - \mathscr{F}_1} \end{pmatrix}$$

and each vector $\mathbf{u} \in \mathbb{Z}^{k_2}$ can be written as

$$\mathbf{u} = (\mathbf{u}_{|\mathscr{F}_1}, \mathbf{u}_{|\mathscr{F}_2 - \mathscr{F}_1}) \qquad \text{with } \; \mathbf{u}_{|\mathscr{F}_1} \in \mathbb{Z}^{k_1} \, .$$

Theorem 1. *If $\mathscr{F}_1$ and $\mathscr{F}_2$ are two fractions with $\mathscr{F}_1 \subset \mathscr{F}_2$, then the circuits in $\mathscr{C}(X_{\mathscr{F}_1})$ are*

$$\{\mathbf{u}_{|\mathscr{F}_1} \ : \ \mathbf{u} \in \mathscr{C}(X_{\mathscr{F}_2}) \ \textit{with} \ \operatorname{supp}(\mathbf{u}) \subset \mathscr{F}_1\}\,.$$

The result above lead us to the construction of an algorithm for finding robust fractions. The strategy is as follows. First, a good fraction should avoid as much as possible the circuits with support on p points or less. Second, small circuits are worse than large circuits, since they are contained in a larger number of minimal fractions, leading to a higher loss in robustness. Thus, at each step a loss function is computed for each point R of the current fraction as the number of minimal fractions becoming non-estimable when removing the point R. In formulae:

$$L(R) = \sum_{\mathbf{u}} \binom{n - \#\operatorname{supp}(\mathbf{u})}{p - \#\operatorname{supp}(\mathbf{u})} \tag{2}$$

where the sum is taken over all the circuits $\mathbf{u}$ in the current fraction containing the point R. Notice that the formula in Eq. (2) does not guarantee that the relevant minimal fractions are all distinct. The formula should be viewed as a first-order approximation of the inclusion-exclusion formula.

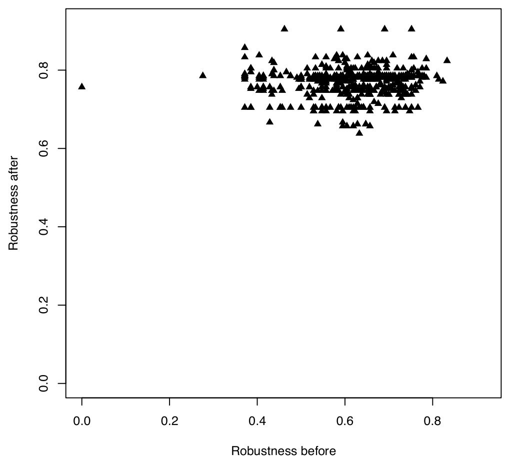
Therefore, we can define a selection algorithm as detailed below. It works like an exchange algorithm as introduced for optimal designs in [10]. To run such an algorithm we only need the circuit basis $\mathscr{C}(X_{\mathscr{D}})$, and we extract the circuits of $\mathscr{C}(X_{\mathscr{D}})$ with support on p points or less. We denote this set of circuits by $\mathscr{C}^p(X_{\mathscr{D}})$.

1. Starting with an arbitrary fraction $\mathscr{F}$ of a specified size n;
2. Repeat:
   a. Consider the circuits of $\mathscr{C}^p(X_{\mathscr{D}})$ which are contained in $\mathscr{F}$;
   b. For each point R in $\mathscr{F}$ compute its associated loss $L(R)$ as the (weighted) number of circuits which include R;
   c. Take all the points with the highest loss and build up all possible pairs with one point not in $\mathscr{F}$. Make the exchange using the pair which reduces as much as possible the number of circuits contained in the fraction.
   d. If no reduction is possible, then break.

3 Examples and Final Remarks

We describe the use of the algorithm on some examples where the candidate sets are full factorial designs. We consider four and five 2-level factors with the main-effect model and two mixed-level cases. In both mixed-level cases we consider three factors, with 2, 3 and 4 levels. In the first case we work with the main-effect model without interactions and in the second one with the main-effect model plus the interaction between the second and the third factor. The candidate sets are the 2^4, the 2^5 and the $2 \times 3 \times 4$ full factorial designs respectively. The algorithm is used for finding robust fractions with different sizes. For each case the algorithm has been used starting from 500 randomly selected fractions. For each case, in Tables 1 and 2 the mean value $\bar{r}_B$ of the robustness of the randomly selected fractions, the mean value $\bar{r}_A$ and the maximum value $\max r_A$ of the robustness of the fractions which are the output of the algorithm are reported. It is worth noting that in all cases the mean of the final robustness ($\bar{r}_A$) is greater than the mean of the robustness of the starting fractions ($\bar{r}_B$), and the efficiency of the algorithm is strong especially in the binary cases. In Fig. 1 the 500 pairs of robustness of the starting fraction and robustness of the final fraction are reported in the case of 5 factors and $n = 10$ as pre-specified size of the fraction. We observe that the distribution of the robustness of the final fractions is better than that of the starting fractions both in terms of mean and dispersion.

Fig. 1 2^5 design and fractions of size $n = 10$. Robustness after *vs* robustness before the algorithm on a set of 500 fractions.



	4 factors			5 factors		
n	$\bar{r}_B$	$\bar{r}_A$	$\max r_A$	$\bar{r}_B$	$\bar{r}_A$	$\max r_A$
8	0.6889	0.8214	0.8214	0.6134	0.8163	0.9643
10	0.6909	0.7619	0.7619	0.6203	0.7700	0.9048
12	0.6887	0.7222	0.7222	0.6164	0.7589	0.8171
14	0.6887	0.6933	0.6933	0.6134	0.7162	0.7586

Table 1 Mean and maximum values of the robustness: 2^4 and 2^5 designs under the main-effect model.

The computation of the circuit basis is actually feasible only for small designs, with at most 6 to 9 factors, depending on the number of interactions included in the model. We are working on an algorithm that uses only the circuits with mini-

	2×3×4 no interaction			2×3×4 with interaction		
n	$\bar{r}_B$	$\bar{r}_A$	$\max r_A$	$\bar{r}_B$	$\bar{r}_A$	$\max r_A$
14	0.3818	0.4482	0.5216	0.0086	0.2857	0.2857
16	0.3793	0.3985	0.4399	0.0097	0.0571	0.0571
18	0.3838	0.3854	0.3986	0.0102	0.0224	0.0224
20	0.3833	0.3853	0.3906	0.0098	0.0132	0.0132

Table 2 Mean and maximum values of the robustness: $2 \times 3 \times 4$ model under the main-effect and the first-order-interaction models.

mal support. This new version of the algorithm could be used also in the case of large designs, since in most cases the circuits with minimal support can be defined theoretically, without any computation.

Acknowledgements

This paper is part of a joint project with Henry Wynn, Emeritus Professor of Statistics, London School of Economics. Roberto Fontana gratefully acknowledges financial support from the Italian Ministry of Education, University and Research (MIUR), “Dipartimenti di Eccellenza” grant 2018-2022.

References

1. Butler, N.A., Ramos, V.M.: Optimal additions to and deletions from two-level orthogonal arrays. J. R. Stat. Soc. Ser. B **69**(1), 51–61 (2007)
2. Fontana, R., Rapallo, F.: On the aberrations of mixed level orthogonal arrays with removed runs. Stat. Pap. (Berl.) **60**(2), 479–493 (2019)
3. Fontana, R., Rapallo, F., Rogantin, M.P.: A characterization of saturated designs for factorial experiments. J. Stat. Plan. Inference **147**, 205–211 (2014)
4. Ghosh, S.: On robustness of designs against incomplete data. Sankhya Ser. B **40**(3/4), 204–208 (1979)
5. Ghosh, S.: Robustness of BIBD against the unavailability of data. J. Stat. Plan. Inference **6**(1), 29–32 (1982)
6. Grayson, D.R., Stillman, M.E.: Macaulay2, a software system for research in algebraic geometry (2019). URL http://www.math.uiuc.edu/Macaulay2/
7. Street, D.J., Bird, E.M.: *D*-optimal orthogonal array minus *t* run designs. J. Stat. Theory Pract. **12**(3), 575–594 (2018)
8. Sturmfels, B.: Gröbner bases and convex polytopes, *University Lecture Series*, vol. 8. American Mathematical Society, Providence, RI (1996)
9. 4ti2 team: 4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces (2018). URL https://4ti2.github.io
10. Wynn, H.P.: The sequential generation of *D*-optimum experimental designs. Ann. Math. Stat. **41**(5), 1655–1664 (1970)
11. Xampeny, R., Grima, P., Tort-Martorell, X.: Which runs to skip in two-level factorial designs when not all can be performed. Qual. Eng. **30**(4), 594–609 (2018)