Annals of the Institute of Statistical Mathematics (2023) 75:683–704
https://doi.org/10.1007/s10463-022-00860-4

Generation of all randomizations using circuits

Elena Pesce[1] · Fabio Rapallo[2] · Eva Riccomagno[3] · Henry P. Wynn[4]

Received: 13 July 2022 / Revised: 1 November 2022 / Accepted: 17 November 2022 /
Published online: 23 December 2022

Abstract
After a rich history in medicine, randomized control trials (RCTs), both simple and complex, are in increasing use in other areas, such as web-based A/B testing and planning and design of decisions. A main objective of RCTs is to be able to measure parameters, and contrasts in particular, while guarding against biases from hidden confounders. After careful definitions of classical entities such as contrasts, an algebraic method based on circuits is introduced which gives a wide choice of randomization schemes.

Keywords Algebraic statistics and combinatorics · A/B testing · Bias and confounders · Big data · Design of experiments

1 Introduction

There are ways in which a regression model can be biased because of the neglect of hidden variables, sometimes called hidden confounders. To some extent these biases can be removed using randomization. A major source of conceptual difficulty is the continuing distinction between passive observation, characterized by the terms

✉ Fabio Rapallo
fabio.rapallo@unige.it

Elena Pesce
elena_pesce@swissre.com

Eva Riccomagno
riccomagno@dima.unige.it

Henry P. Wynn
h.wynn@lse.ac.uk

[1] Swiss Re Institute, Swiss Re Management Ltd, Mythenquai 50/60, 8022 Zurich, Switzerland

[2] Department of Economics, Università di Genova, Via F. Vivaldi 5, 16126 Genoa, Italy

[3] Department of Mathematics, Università di Genova, Via Dodecaneso 35, 16146 Genoa, Italy

[4] London School of Economics, London, WC2A 2AE, UK

"observational study" and controlled experiment. In addition this distinction is flavored by different intellectual traditions. In most fields a controlled experimental design is conceived as an *intervention*. Thus one talks about setting the level of a variable X, or applying a treatment or treatment combination. Rather than interfere too much with the state of Nature one may simply select a value of X which is already in a population, such as selecting a subject of a particular age. Stratification is in this category as is "matching", observing (or treating) a collection of subjects who are close in terms of some multivariate metric applied to the possible confounders. "Natural Experiments" exploit opportunities where Nature has unwittingly designed an experiment for us. For a very thorough compendium of experimental design methodology both as intervention and as selection, see Dean et al. (2015).

Traditions in agriculture and socio-medical sciences have stressed the role of randomization, and indeed the method has been described as one of the greatest contributions of statistics to scientific methodology; a major review is Cox (2009) which goes a long way toward updating earlier discussions, such as Kempthorne (1955). After a long period in which factorial and optimum controlled experiments may be seen to have had a dominant role, influenced by success in product design and quality improvement, randomization is making a come back, if indeed it ever left the limelight. It is now used extensively outside its traditional areas of clinical trials under the generic term randomized control trials, RCT. Notably, there is a fast growing application to experiments in social media, under the heading A/B testing in on-line marketing, see Kohavi and Longbotham (2017), and to socio-technical experiments, such as smart metering in homes and transport, see e.g., Guzowski et al. (2014). Other important developments are in the field of "big data", where data are often collected without experimental design being used at all, so that biases can be a serious impediment to model building, see Drovandi et al. (2017), Pesce et al. (2019, 2022).

There seems to be no doubt that in nearly all fields the removal of biases in modeling is a major reason to randomize. The question then remains as to whether the randomization, or rather the randomization distribution, is to be used in the analysis, e.g., probability statements are made based on the randomization, for example, using nonparametric tests, or whether randomization should only be used in the design, e.g., for bias reduction. The latter approach is probably more common and is adopted here. A compromise position is a minimax approach which is closely related to the use of randomization in finite population sampling, see Scott and Smith (1975), Stenger (1979), Stigler (1969), Wynn (1977).

Our approach can be considered a contribution to the subtle relationship between randomization and combinatorial design, see Bailey and Rowley (1987). It is based on the theory of *circuits*, which are already studied in operations research (Simões Pereira 1975) and algebraic statistics (Fontana et al. 2022). The better known extensions of simple RCT such as block randomization, stratified randomization and the less covered hierarchical randomization are covered by our methods, and we shall return to this claim in the last section.

After a straightforward formulation of the problem, we formally define *valid* randomization schemes in Sect. 4, followed by a short discussion on analysis in Sect. 5. Sections 6 and 7 are the main developments, with Sect. 6 describing a sufficient

condition under which unions of non-negative binary circuits give a valid randomization. Section 7 gives some special conditions. Final considerations in Sect. 8 conclude the paper.

1.1 A/B testing

Some of the disparate interpretations of randomization can be understood from a simple A/B testing (RCT) experiment, which is typically used to assess the difference between the effect of two treatments A and B with effect parameters θ_A and θ_B, respectively. That is, we want to estimate $\phi = \theta_A - \theta_B$.

A standard model for a response variable Y is to write for subjects i and j receiving treatments A and B, respectively

$$\begin{aligned} Y_{Ai} &= \theta_A + \delta_{Ai},\ i = 1, \ldots, n_A, \\ Y_{Bj} &= \theta_B + \delta_{Bj},\ j = 1, \ldots, n_B \end{aligned}$$

where n_A, n_B are the respective sample sizes and δ_{Ai}, δ_{Bj} are unit effects of other influences, be they errors of measurement or other (hidden) factors effects. Y_{Ai} and Y_{Bj} are therefore specializations of Y for the two sub-populations A and B. The naive estimate of the treatment difference is

$$\hat{\phi} = \hat{\theta}_A - \hat{\theta}_B\,.$$

Here the estimates of θ_A and θ_B are given by the respective sample means:

$$\hat{\theta}_A = \bar{Y}_{A\cdot}\,,\ \hat{\theta}_B = \bar{Y}_{B\cdot}\,,$$

where for instance $\bar{Y}_{A\cdot}$ is the usual notation for the average of measurements over group A. The standard argument, and this is probably also the common sense argument of non-experts, is that if we randomize then the difference between the mean values of the deviations due to other factors will cancel out: $\delta_{A\cdot} - \delta_{B\cdot}$, will be approximately zero and will not perturb $\hat{\phi}$. Of course, if δ_{Ai}, δ_{Bj} are random with standard assumptions then $\hat{\phi}$ is both the least squares estimate and the best linear unbiased estimate of ϕ.

A critical question is: what does the model mean, both scientifically and predictively? What are θ_A, θ_B and ϕ? In other words, do parameter values refer to the finite population from which the sample was taken or to which the treatment were applied? Or is there some larger population of which the population of units under study is a subpopulation, such as all present and future subjects who may benefit from a vaccination decision based on the results of the experiment? Or, are A and B a "crucial experiment", to decide between two scientific theories? These questions are important also with the A/B testing experiments on people using social media. The commercial opportunities in terms of the use of huge (big) data sets come with a risk of bias arising from any number of demographic and operations factors. It is almost impossible to describe the population of social media users but if bias can be removed in some simple way then the estimates can genuinely reflect peoples' choices and behavior.

A naive but rather universal conclusion is something like: after randomization we can use the model. This is expressed as part of expert advice: make sure you randomize your blocks. On the one hand this paper takes this simple approach, but on the other introduces a special technique, based on circuits, to decompose an experiment into mutually exclusive blocks in each of which randomization can be carried out separately. Some solutions comprise recognizable combinatorial designs, such as matching and stratification. All the others can be derived from the circuits which can be computed from running the program `4ti2` (4ti2 team 2018). Our approach provides a full solution to the problem of block randomization to control bias, up to the computational feasibility, see Sect. 6.1.

2 Contrasts

Consider an experiment giving a random sample $Y_1, \ldots, Y_n$ and the following:

Definition 1 A linear function $Z = \sum_{i=1}^{n} c_i Y_i$ with fixed coefficients $\{c_i\}$ is called an *empirical contrast* if $\sum_{i=1}^{n} c_i = 0$.

In the A/B case randomization is particularly suited to situations in which standard estimates are unaffected by a uniform shift of the observations, which is then subtracted out.

Now consider a standard regression model in the form

$$Y(x) = \sum_{j=1}^{p} \theta_j f_j(x) + \epsilon,$$

for functions $\{f_j(x)\}$, x a generic point in some design space $\mathcal{X}$, for parameters $\{\theta_j\}$, and ϵ a random error with the usual assumptions (zero mean and constant variance).

An experimental design $D = \{x^{(i)},\ i = 1, \ldots, n\}$, with sample size $|D| = n$, has design matrix

$$X = \{f_j(x^{(i)})\}$$

with dimension $n \times p$, and we express the standard regression set-up by:

$$\mu = \mathbb{E}(Y) = X\theta,$$

where θ is a vector of parameters with length p and $\mathbb{E}$ is the expectation. Definition 2 follows standard terminology in regression models and design of experiments, see Das and Jain (1970).

Definition 2 For a standard regression model a *parametric contrast* is defined as the expectation of an empirical contrast.

In the following we exemplify the basic idea to divide experiment into disjoint blocks in each of which we randomize, and then combine the results.

Example 1 (2^2 experiment) We consider a simple example from linear regression, namely a 2^2 factorial design problem, with ± 1 levels and no replication (for simplicity). We take the model without an interaction

$$\mathbb{E}(Y) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \,,$$

so that design matrix is

$$X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix}.$$

If we randomize a large population and uniformly apply the four combination of the design, $\{\pm 1, \pm 1\}$, the potential bias effect will be negligibly small because the estimators of the θ-parameters are unbiased.

But there is an alternative. Split the population into two groups, randomize each separately and apply the controls $(x_1, x_2) = \{(1,1), (-1,-1)\}$ to the first group and $\{(1,-1), (-1,1)\}$ to the second group. Then we can estimate $\theta_1 + \theta_2$ from the first group and $\theta_1 - \theta_2$ from the second group. Combining these estimates gives the same result as if we randomized over the whole 2^2 experiment. Note that the parameters θ_1 and θ_2 and their estimates are already respectively parametric contrasts and empirical contrasts, with contrast coefficients equal to $c = \frac{1}{4}(1,-1,1,-1)^T$ and $c = \frac{1}{4}(1,-1,-1,1)^T$, respectively. This can be seen as splitting the 2^3 experiment into two (randomized) A/B experiments.

3 Writing a model in contrast form

In the case of the orthogonal design described above the X-matrix takes the form

$$X = [\xi \,:\, X_1],$$

where ξ is the n-vector of ones, for the constant (intercept) term, and X_1 is a matrix with dimension $n \times (p-1)$ orthogonal to ξ, that is $\xi^T X_1 = 0$. We describe such an X-matrix as being *in contrast form*. All empirical and parametric contrasts are derived from X_1. Thus we can prove the following lemma.

Lemma 1 *For a regression model with $\mu = \mathbb{E}(Y) = \mathbb{E}(\tilde{X}\theta)$, written in contrast form $\tilde{X} = [\xi \,:\, X_1]$ the set of all parametric contrasts is $\{c^T\mu \,:\, c^T X_1 = 0$ and $\xi^T c = 0\}$.*

Proof This follow since $\mathbb{E}(c^T Y) = c^T[\xi \,:\, X_1]\theta = (c^T\xi, c^T X_1)\theta$. □

Notice that from any model with integer design matrix X it is always possible to derive a reparametrization with design matrix $\tilde{X}$ written in contrast form. In Sect. 6, we will use the vector ξ and we will exploit its orthogonality to X_1 to

study the connections between randomizations and circuits. Note that the assumption of integer design matrix X is made here only to simplify the computation of the circuits introduced in Sect. 6. The theory here is valid for design matrices with rational entries. A design matrix with rational entries can be multiplied by a constant, namely the least common multiple of the denominators, to obtain a matrix with integer entries whose columns generate the same vector space.

Lemma 2 *Every model $Y = X\theta + \epsilon$ including the intercept can be transformed to contrast form as $Y = \tilde{X}\phi + \epsilon$, where $\tilde{X} = [\xi : X_1]$ is full-rank and has the same column space as X and $\xi^T X_1 = 0$.*

Proof We can easily determine the reparametrization which the transformation requires. Starting with:

$$\tilde{X}\phi = X\theta,$$

we simply solve for ϕ:

$$\phi = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T X\theta.$$

□

From Lemma 2, a design matrix X with column space containing the vector $\xi = (1, 1, \ldots, 1)^T$ can be transformed to contrast form. The term contrast is especially prevalent in Analysis of Variance (ANOVA) models, that is additive models for qualitative factors in which each level of each factor provides a parameter. The classical notation for a two-way $I \times J$ table with two factors is that the additive model would have parameters $\alpha_i, (i = 1, \ldots, I)$ and $\beta_j, (j = 1, \ldots, J)$ and the model for the observations Y_{ij} is

$$Y_{ij} = \alpha_i + \beta_j + \epsilon_{ij}\,, \tag{1}$$

where $\{\epsilon_{ij}\}$ are the random errors with standard assumptions. We show below how the reparametrization to obtain a model in contrast form works with an example for a model as in Eq. 1.

Example 2 Let $I = J = 2$. By using indicator variables and setting $\theta = (\alpha_1, \alpha_1, \beta_1, \beta_2)^T$ we write the model in regression form, $\mathbb{E}(Y) = X\theta$ where

$$X = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

This X-matrix is *not* in contrast form, but it can be transformed into contrast form:

$$\tilde{X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix}.$$

From this, the reparametrization is:

$$\begin{aligned} \phi_0 &= \frac{1}{2}(\alpha_1 + \alpha_2 + \beta_1 + \beta_2), \\ \phi_1 &= \frac{1}{2}(\alpha_1 - \alpha_2), \\ \phi_2 &= \frac{1}{2}(\beta_1 - \beta_2). \end{aligned}$$

We have limited the analysis to the decomposition of $\tilde{X}$ into $[\xi : X_1]$ since for randomization we are interested in the decomposition of the vector ξ, but the results in this section and many results about the circuit bases in the next sections could be written in general for a decomposition of $\tilde{X}$ into $[X_2 : X_1]$ with $X_2^T X_1 = 0$.

Note that when a full-rank matrix $\tilde{X}$ is decomposed into $[\xi : X_1]$, also the matrix X_1 is full-rank. To avoid trivialities, we also assume that all the rows of the matrix X_1 are not null, i.e., each design point is involved in at least one contrast.

4 Valid randomizations

Using the representation of the design matrix in contrast form we can provide a catalogue of valid randomization systems to address the question stated earlier in Sect. 1.1 in the framework of A/B experiments. The elements of the catalogue can be computed and in the case of unimodular X_1 matrix (see Sect. 7) this catalogue is complete. The separation into blocks is a partition of the observations so that there are at least two observations in each element of the partition, as described by the following definitions giving the only relevant randomizations to study contrasts.

Definition 3 For observations Y_i, $(i = 1, \ldots, n)$ a *potential randomization system R* is a set partition of $\mathcal{N} = \{1, 2, \ldots, n\}$, namely a decomposition of $\mathcal{N}$ into disjoint exhaustive subsets, $R_1, \ldots, R_k$, called blocks, of size 2 or more:

1. $\bigcup_{i=1}^{k} R_i = \mathcal{N}$
2. $R_i \cap R_j = \emptyset$, $1 \le i < j \le k$
3. $|R_i| \ge 2, i = 1, \ldots, k$

Definition 4 For a regression model and experimental design D_n with sample size n, a design matrix in contrast form $[\xi : X_1]$ and a potential randomization system $\{R_1, \ldots, R_k\}$, let $z^{(i)} = (z_{i,1}, \ldots, z_{i,n})$ the binary vectors defined by

$$z_{i,j} = \begin{cases} 1, & i \in R_j \\ 0, & i \in \mathcal{N} \setminus R_j \end{cases}.$$

The potential randomization system is a *valid randomization system* if $z^{(i)}$ is orthogonal to X_1, i.e., $(z^{(i)})^T X_1 = 0$, for all $i = 1, \ldots, k$.

The case where $R = \mathcal{N}$, we refer to as full randomization. The next two examples are familiar in the sense that the orthogonal blocks are easily associated with addition factors or parameters in an orthogonal design. The third example may be less familiar.

4.1 Factorial fractions

We consider a 2^3 factorial experiment for main effects. The standard X-matrix is already in contrast form:

$$X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix}.$$

In addition to a full randomization, represented by $\{1, 2, 3, 4, 5, 6, 7, 8\}$, there are two different randomization systems and we list the R_j partitions for each:

1. $\{1, 4, 6, 7\}, \{2, 3, 5, 8\}$;
2. $\{1, 8\}, \{2, 7\}, \{3, 6\}, \{4, 5\}$.

These two distinct randomizations of this example correspond to familiar decomposition into blocks based on abelian groups (see e.g., Box et al., 1978). The first arrives from a 2^{3-1} experiment with defining contrast subgroup in classical notation

$$I = ABC.$$

The second corresponds to the 2^{3-2} with subgroup

$$I = AB = BC = AC.$$

For those more familiar with the algebraic design of experiments, these solutions are the point ideal corresponding respectively to the solutions of

$$(1) : x_1 x_2 x_3 = \pm 1, \quad \text{and} \quad (2) : (x_1 x_2, x_2 x_3) = (\pm 1, \pm 1).$$

4.2 Tables and Latin squares

Consider an $I \times I$ table with the usual additive model. A Latin square based on the table has the usual definition. If $I = 3$ there are two mutually orthogonal Latin squares; in traditional notation:

$$\begin{array}{ccc|ccc} A & B & C & a & b & c \\ C & A & B & b & c & a \\ B & C & A & c & a & b \end{array}$$

Each square gives a different valid randomization based on the letters. Labeling the observations left-to-right and top-to-bottom the respective blocks are (ignoring commas)

$$\{159\}, \{267\}, \{348\}, \quad \{168\}, \{249\}, \{357\}.$$

We state the general result without proof and in the terminology of this example.

Lemma 3 *For an $I \times I$ additive Analysis of Variance model a set of mutually orthogonal Latin squares provides a set of alternative valid randomizations.*

4.3 *k*-out-of-2*k* choice experiments

Choice experiments are those in which subjects are asked to score a selection of attributes from a portfolio of attributes. Models are fitted to experimental data in an effort to discover subjects' (hidden) preference order.

Suppose there are $n = 4$ attributes and each subject is offered $k = 2$ attributes, labeled 1, 2, 3, 4. There are six selection pairs

$$\{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}.$$

An additive preference model has (without replication) the six values $Y_{i,j}$ with the model

$$Y_{ij} = \alpha_i + \alpha_j + \epsilon_{i,j} \quad (i,j = 1,2,3,4; i < j).$$

We are interested in contrast $\alpha_i - \alpha_j$, because their estimates would yield an estimated preference order. In this case:

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

This gives a choice of X_1:

$$X_1^T = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \end{bmatrix},$$

and the randomization: $\{1,6\},\{2,5\},\{3,4\}$, where the integers refer to selection pairs.

5 Analysis

The informal approaches we have taken is that, for large samples randomization has approximately the effect of introducing a block parameter. Our condition of orthogonality in the definition of valid randomization and as exemplified, has so far ignored the fact that in standard terminology blocks do not have to be orthogonal. Indeed, there is rich theory of balanced incomplete blocks (BIBD) both from combinatorial and from optimal design theory. We note here some basic facts about orthogonal versus non-orthogonal blocks.

1. For orthogonal designs we set up a model in which every binary vector orthogonal to the X_1 matrix is allocated a block parameter, then only under orthogonality is the usual Least Square Estimate (LSE) of the θ-parameters the best and there is no bias of these estimates from the block effects.
2. In the non-orthogonal blocks design case, if we use the LSE of the θ-parameters assuming that the block parameters are zero, when they are not, then the block parameters introduce bias.
3. In the non-orthogonal blocks case the "proper" LSE estimate of the θ-parameters in the presence of the block parameters, will be unbiased but will have higher variances than in case (2) above. This can be expressed by the Loewner ordering: one covariance matrix is "smaller" than the other if the difference is non-negative definite.

Models with non-orthogonal blocks with a specified block effect, require some effort to model or at least interpret the block affect, for example the effect of day if the experiment is conducted over days. In such cases a bias model is required. But where bias is caused by hidden, unspecified, confounders, such a bias model seems somewhat artificial. The effects are too artificial to model but sufficiently present that we prefer orthogonality.

6 Circuit basis for randomization

In this section, we introduce the notion of circuits of a matrix which allows a novel approach to the problem of randomization. The proposed analysis, based on tools from Algebraic Statistics, leads to the enumeration of all possible randomization schemes. In this setup a randomization is given by the decomposition of the vector $\xi = (1,\dots,1)^T$ into binary vectors:

$$\xi = \xi_1 + \cdots + \xi_k \tag{2}$$

where each vector $\xi_h \in \{0,1\}^n$ satisfies $\xi_h^T X_1 = 0$, $h = 1, \ldots, k$. Such binary vectors ξ_h are called binary randomization vectors. Next, we introduce the circuits and their main properties. When all randomization vectors cannot be decomposed into binary vectors with smaller support we have a non-decomposable randomization.

Definition 5 Given a randomization of ξ into binary vectors as in Eq. (2), the vector ξ_h is a non-decomposable randomization vector if there is no decomposition $\xi_h = \xi_{h,1} + \xi_{h,2}$ with $\xi_{h,1}^T X_1 = 0$ and $\xi_{h,2}^T X_1 = 0$. If all the vectors $\xi_1, \ldots, \xi_k$ are non-decomposable, Eq. (2) defines a non-decomposable randomization.

Let A be an integer-valued matrix with d rows and n columns. For our purposes, we can assume that $A = X_1^T$. Let $u \in \mathbb{Z}^n$ be an integer-valued vector, u^+ be the positive part of u, namely $u_i^+ = \max(u_i, 0)$, $i = 1, \ldots, n$, and u^- be the negative part of u, namely $u_i^- = -\min(u_i, 0)$, $i = 1, \ldots, n$, so that $u = u^+ - u^-$. Moreover, denote with $\mathrm{supp}(u)$ the support of u, i.e.,

$$\mathrm{supp}(u) = \{i \in \{1, \ldots, n\} \ : \ u_i \neq 0\}.$$

Definition 6 A *circuit* of A is an integer-valued vector u in $\ker(A)$, i.e., $Au = 0$, with the following minimality properties:

1. u has minimal support, i.e., there is no other circuit v with $\mathrm{supp}(v) \subset \mathrm{supp}(u)$.
2. u is irreducible: if v is an integer-valued vector in $\ker(A)$ with $\mathrm{supp}(v) = \mathrm{supp}(u)$, then $v = ku$ for some $k \in \mathbb{N}$.

Definition 7 The set of all circuits of the matrix A is named the *circuit basis* of A and is denoted with $\mathcal{C}(A)$.

The circuit basis $\mathcal{C}(A)$ is always finite. The minimal support property gives rise to a number of interesting properties of $\mathcal{C}(A)$. We recap in the following proposition the special features of the circuits we will use for describing randomization. For the proofs and further details the reader can refer to Sturmfels (1996).

Proposition 1 *Let A be an integer-valued matrix with dimensions $d \times n$ and suppose that* $\mathrm{rank}(A) = d$.

1. *The circuit basis $\mathcal{C}(A)$ is subset compatible, i.e., for a sub-matrix A' with $n' < n$ columns of A, the circuit basis of A' is given by the circuits in $\mathcal{C}(A)$ whose support is contained in the n' columns.*
2. *The cardinality of the support of a circuit in $\mathcal{C}(A)$ is at most $d + 1$.*
3. *Each vector v of* $\ker(A)$ *can be written as rational non-negative linear combination of circuits, i.e.,*

$$v = \sum_{h=1}^{n-d} q_h u_h \,,\; q_h \in \mathbb{Q}_+$$

and the u_h are conformal with v.

The term "conformal" in Item (3) of Proposition 1 means that $\text{supp}(u_h^+) \subset \text{supp}(v^+)$ and $\text{supp}(u_h^-) \subset \text{supp}(v^-)$.

The first key observations for randomization follow directly from the fact that a circuit lies in $\ker(A)$.

Lemma 4 *Any non-negative binary circuit of $A = X_1^T$ provides a randomization vector.*

Proof When a non-negative binary circuit ξ_1 gives a valid randomization, then also $\xi_2 = \xi - \xi_1$ is a binary non-negative vector in $\ker(A)$ so that the decomposition $\xi = \xi_1 + \xi_2$ is a valid randomization. □

Note that the vector ξ_2 in the proof may be a circuit itself (and in such a case we call $\xi = \xi_1 + \xi_2$ a non-decomposable randomization), or not. In the latter case, the vector ξ_2 can be decomposed into the sum of non-negative circuits by virtue of Proposition 1, Item (3).

The decomposition of ξ in Eq. (2) and the argument above show that valid randomizations generate a lattice, partially ordered by set inclusion, indeed: (1) circuits sit at the most refined level of the lattice and (2) less refined randomization schemes are obtained by merging two lattice elements into their join. This connection with lattice (and matroids) is taken up again in the discussion section. From Proposition 1, Item (3), and Lemma 4, we see that the circuit basis, and in particular the set of non-negative circuits, is the natural tool to find valid non-decomposable randomizations. In general, if the vector ξ can be written as the sum of binary non-negative circuits we have a valid randomization. The main problem posed in this paper is to provide conditions for when the converse holds, that is to provide classes of experimental designs for which every randomization vector ξ_h is a circuit. In the next section we will describe an important class, here we have a useful sufficient condition.

Lemma 5 *If ξ_1 is a non-negative binary randomization vector with two nonzero elements* ($\#\text{supp}(\xi_1^+) = 2$), *then it is a circuit of X_1^T.*

Proof In view of Proposition 1, Item 3, it is enough to prove that there is no circuit ξ_1 with exactly one nonzero element. By contradiction, suppose that such a vector ξ_1 exists and, without loss of generality, suppose that $\xi_1 = (1, 0, \ldots, 0)$. Since ξ_1 is in the kernel of X_1^T, we have $X_1^T \xi_1 = 0$ and this implies that the first column of X_1^T is a column of zeros. This is in contradiction with the fact that all rows of X_1 are not zero. □

Thus, for every ξ_1-vector in example covered by Lemma 5, there are two rows of X_1^T which have opposite signs. This is the case in Sect. 4.3 which yields the following result.

Corollary 1 *Any k-out-of-2k choice experiment is a valid randomization with blocks of size 2.*

Proof For a k-out-of-$2k$ experiment we can construct an X matrix with rows corresponding to k-tuples and the rows in lexicographic order. Then as for the example in Sect. 4.3 we pair them: the first with the last, second with the second to last and so on, assigning -1 and $+1$ respectively, to construct the X_1 matrix. Let $n = \binom{2k}{k}$, then the valid randomization blocks are the selection pairs:

$$\{1, n\}, \{2, n-1\}, \ldots, \{n/2, n/2-1\},$$

which follows because X_1 is of the type discussed in Lemma 5. □

This shows that a valid randomization with binary vectors each with two nonzero binary vectors can be found by inspecting the list of all circuits.

6.1 Computation of circuits

To find the randomization systems from the circuit basis, we start from the design matrix X, we write it in contrast form $\tilde{X}$, and we extract the contrast matrix X_1 as described above. The actual computation of the circuits of the matrix X_1 can be done with the software package `4ti2`, see 4ti2 team (2018). In `4ti2` there is a function called `circuits` to compute the circuits of an integer matrix. The algorithms to compute circuits in `4ti2` belong to the class of combinatorial algorithms, and thus there is a limitation on the size of the matrices for which the computation of the circuit is actually feasible. In our experiments, problems with a set of points up to 50 are easily processed, but the execution time increases fast with the number of points. However, all the contrast matrices illustrated in this paper have been processed by `4ti2` in less than 0.1 seconds. `4ti2` is now available also within the symbolic software `Macaulay2`, see Grayson and Stillman (2019), and there are `R` packages available which allow the communication between `R` and `Macaulay2`, leading to a flexible use of the symbolic computations into statistical analysis, see Kahle et al. (2020).

Example 3 Using the function `circuits` for the contrast matrix of the 3-out-of-6 problem, we obtain three circuits as expected

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Example 4 Computing the circuits for the 2^3 design with contrasts on the main effects, we obtain the circuits described in the previous sections. The `4ti2` output consists of 20 circuits, 6 of which are non-negative:

$$\begin{bmatrix} 0&0&0&1&1&0&0&0 \\ 0&0&1&0&0&1&0&0 \\ 0&1&0&0&0&0&1&0 \\ 0&1&1&0&1&0&0&1 \\ 1&0&0&0&0&0&0&1 \\ 1&0&0&1&0&1&1&0 \end{bmatrix}.$$

This yields the two randomization schemes

$$\{1,4,6,7\},\{2,3,5,8\} \qquad \text{and} \qquad \{1,8\},\{2,7\},\{3,6\},\{4,5\}$$

already discussed. Here, there is only one valid randomization based on 2-ers and only one valid randomization based on 4-ers. (The term *n*-er is a colloquial term for an entity of size *n*.)

With the aid of the circuits we are able to analyze also more complex models where the number of randomization systems is relatively large.

Example 5 In the case of 2^4 design with contrasts on the main effects, the contrast matrix is:

$$X_1^T = \begin{bmatrix} 1&1&1&1&1&1&1&1&-1&-1&-1&-1&-1&-1&-1&-1 \\ 1&1&1&1&-1&-1&-1&-1&1&1&1&1&-1&-1&-1&-1 \\ 1&1&-1&-1&1&1&-1&-1&1&1&-1&-1&1&1&-1&-1 \\ 1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1&1&-1 \end{bmatrix},$$

and the situation becomes more complex. Although 0.02 seconds are enough to obtain the whole set of 456 circuits, the non-negative circuits are now 48 but there are also non-binary circuits with entries equal to 2. Selecting the binary circuits reduces to 32 circuits: 8 circuits with support on two points give a unique randomization based on 2-ers; with the remaining 24 circuits on 4 points we can construct 30 valid randomizations. Each circuit on 4 points is used in 5 possible randomizations. For instance with the circuit

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

one can define 5 randomizations, reported in Fig. 1.

With a large choice of randomization schemes the problem arises as to which to choose. This is discussed briefly in Sect. 8.

Fig. 1 The 5 randomizations for the 2^4 configuration in Example 5 containing the circuit $c = (0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0)$

0	0	0	0	0	1	1	0	1	0	0	1	0	0	0	0
0	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	1
1	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0
0	0	0	0	0	1	1	0	1	0	0	1	0	0	0	0
0	0	0	1	1	0	0	0	0	0	1	0	0	1	0	0
0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	1
1	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0
0	0	0	0	0	1	1	0	1	0	0	1	0	0	0	0
0	0	0	1	1	0	0	0	0	1	0	0	0	0	1	0
0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	1
1	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0
0	0	0	0	0	1	1	0	1	0	0	1	0	0	0	0
0	0	1	0	0	0	0	1	0	1	0	0	1	0	0	0
0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	1
1	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0
0	0	0	0	0	1	1	0	1	0	0	1	0	0	0	0
0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	1
0	1	0	0	0	0	0	1	0	0	1	0	1	0	0	0
1	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0

7 Totally unimodular X_1

Although the factorial design and Latin square examples can be considered well-known, because of their orthogonality properties, example in Sect. 4.3 may be less so. So we may ask what is the property of X_1^T for which the full valid randomization system can be found as a set of circuits.

Definition 8 A *totally unimodular matrix A* is one for which all square sub-matrices (including itself if square) have determinant 0, 1, or −1.

Theorem 1 *Let $A = X_1^T$ be the design/model matrix of regression model in contrast form and suppose A is totally unimodular. Then every valid randomization is based on circuits.*

The proof is in two parts. First we need the following lemma, whose proof is based on some technical results from the algebraic theory of toric ideals and Gröbner bases. In order to maintain the focus on the problem of randomization, we do not recall here all the formal definitions of the objects needed in the proof, for which the reader can refer to, e.g., Sturmfels (1996).

Lemma 6 *For a totally unimodular matrix A all non-negative circuit vectors are binary.*

Proof In this statement, the circuits should be seen as represented by the so-called binomials, that is for each circuit $u = u^+ - u^-$ we consider n "dummy" variables $x_1, \ldots, x_n$ and the binomial associated to u is defined as:

$$x^{u^+} - x^{u^-}.$$

These binomials generate a toric ideal $I(A)$. This ideal is very widely studied, for example in algebraic statistics it is the starting point for Markov Chain Monte Carlo simulation for testing hypotheses on multinomial contingency tables, see Diaconis and Sturmfels (1998).

Now, if A is totally unimodular then it is known that the initial ideal $\text{in}(I(A))$ is generated by square-free binomials for any given term-order (required to define a Gröbner basis), see Sturmfels (1996). The initial ideal $\text{in}(I(A))$ of the ideal $I(A)$ is the ideal generated by the leading terms of the polynomials in $I(A)$. Thus, all the binomials in the Universal Gröbner basis $\mathcal{U}(I(A))$ have square-free leading terms.

Finally, the non-negative circuits are elements of $\mathcal{U}(I(A))$, viewed as binomials of the form $x^u - 1$. The leading term is always x^u, it is square-free, and therefore u is binary. □

To complete the proof of Theorem 1 we also need the following result.

Lemma 7 *If the contrast matrix $A = X_1^T$ in a regression model is totally unimodular then every non-decomposable randomization vector ξ is a circuit.*

Proof This is by contradiction. Let ξ_1 be a (non-negative binary) non-decomposable randomization vector and suppose it is not a circuit. Since $\xi_1 \in \ker(A)$, by Proposition 1, Item 3, $\xi_1$has a representation as a non-negative linear combination of circuits $u_1 + \cdots + u_k$. Take one of such circuits u_h. Its support is strictly contained in $\text{supp}(\xi_1)$ and note that $\#\text{supp}(\xi_1) - \#\text{supp}(u_h) > 1$, because ξ_1 is not a circuit and there are no circuits with support on one point. Moreover, the circuit u_h is binary by Lemma 6. So there is a refinement given by $\xi_1 = u_h + (\xi_1 - u_h)$, which contradicts ξ_1 being non-decomposable. □

Proof *(of Theorem 1)* Let

$$\xi = \xi_1 + \cdots + \xi_k\,,$$

be a valid randomization. If it is non-decomposable, then the vectors $\xi_1, \ldots, \xi_k$ are circuits by Lemma 7. If the randomization is decomposable, each vector ξ_h can be decomposed into the sum of non-negative circuits, by Proposition 1, Item 3. By Lemma 6 such circuits are binary and they form a non-decomposable randomization. □

The best known example of a totally unimodular matrix is generated by a directed graph $G(V, E)$. The rows are indexed by vertices and the columns by directed edges with the following rule for entries: if the edge is $e = (i \to j)$ then entries $A_{i,e} = 1, A_{j,e} = -1$ and all other entries in column e are zero. For A to be an X_1 matrix we need it to be (row) orthogonal to $\xi = (1, 1, \ldots, 1)^T$, this requires that for any vertex the number of in-arrows and the number of out-arrows must be the same.

Example 6 Let $|V| = 5, |E| = 15$ and the directed edges (leaving out commas):

$$12, 13, 14, 23, 24, 25, 34, 35, 31, 45, 41, 42, 51, 52, 53.$$

In this example $A = X_1^T$ is

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 \\ 0 & -1 & 0 & -1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

The graph for this example is pictured in Fig. 2. For the X_1 matrix above, there are 33 non-negative circuits from a total of 198 circuits: 5 2-ers, 10 3-ers, 10 4-ers, and 8 5-ers. The valid randomizations we obtained from those circuits are reported in the following table giving the cardinality of the subsets and number r of different choices, classified by the corresponding integer partition.

Randomization	r
$5+5+5$	1
$5+5+3+2$	5
$5+3+3+2+2$	5
$5+2+2+2+2+2$	1
$4+4+3+2+2$	10
$4+3+2+2+2+2$	5
$3+3+3+2+2+2$	5

By the properties of the circuits we know that no proper subset is possible in the previous randomization, so for instance we know that no randomization of the form $5+5+3+2$ can share two 5-ers with the randomization $5+5+5$. However, the $5+5+5$ shares a 5-ers with the randomization $5+2+2+2+2+2$, as shown in Fig. 3.

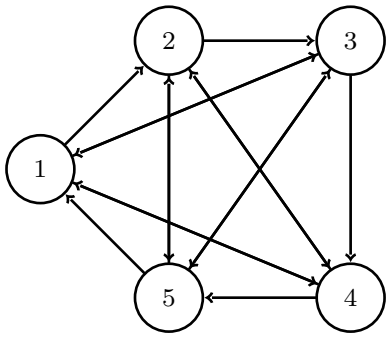

Fig. 2 The directed graph on 5 points in Example 6

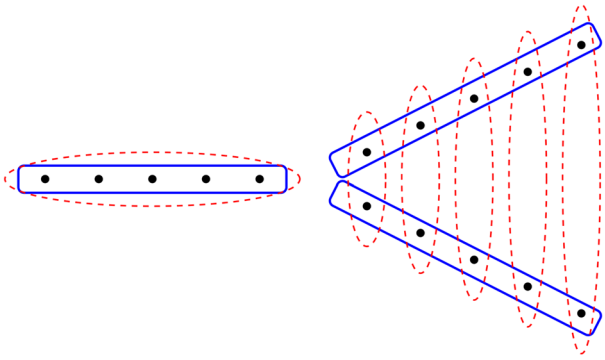

Fig. 3 Two randomizations for the directed graph on 5 points in Fig. 2: a $5+5+5$ randomization (solid lines) and a $5+2+2+2+2+2$ randomization sharing a 5-er (dashed lines)

Example 7 Our final example exploits the existing structure of the design/model environment to make finding the circuits more straightforward, as we saw for factorial designs. The full saturated X-matrix below is taken from a Haar wavelet model on $[-1, 1]$ with depth three from the constant term:

$$X = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & -1 & 0 & -1 & 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & -1 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

Making use of the intrinsic orthogonality we use columns 2,3,4 for the X_1 matrix leaving the last four columns to extract the circuits. Computing the circuit basis for X_1 we obtain 16 circuits with 0-1 entries and with support on 4 points. Each vector has a complementary vector (interchanging the ones and zeros) which together form a randomization scheme with 2 randomized blocks, i.e., of the form $4+4$. This example is small enough that also a direct computation is possible. As an exercise, we find that the circuits can be computed by brute force solving the equations:

$$\begin{aligned} & x_i(1-x_i) = 0, \quad i = 1, \ldots, 8 \\ & x_1 + x_2 + x_3 + x_4 - x_5 - x_6 - x_7 - x_8 = 0 \\ & x_1 + x_2 - x_3 - x_4 = 0 \\ & x_5 + x_6 - x_7 - x_8 = 0 \end{aligned}$$

There are 18 solutions. Excluding the null vector and the vector with all entries equal to 1 (full randomization), we obtain the non-trivial solutions, i.e., 16 binary vectors $(x_1, \ldots, x_n)$ with 4 ones and 4 zeros, which correspond exactly with the 16

circuits computed by `4ti2`. We give just one example of valid randomization for this example to save space. Two non-trivial solutions are

$$(1,0,0,1,1,0,0,1),(0,1,1,0,0,1,1,0)$$

which are confirmed be orthogonal to the model columns 2,3,4. They correspond to the randomization

$$\{1,4,5,8\},\{2,3,6,7\}\,.$$

Finally, we briefly discuss the unimodularity assumption. Although unimodularity seems to be a restrictive assumption, a number of models in important applications are defined by a unimodular matrix. For instance the independence model for two-way tables has an unimodular design matrix, the Kronecker product of two unimodular matrices is unimodular, providing a large class of models with unimodular design matrix. Other examples comes from optimization and graph theory, thus for statistical network models. The coefficient matrix of the constraints in the linear programming formulation of the maximum flow problem is unimodular. An example from graph theory has been used in Example 6.

There are criteria to check whether a matrix is totally unimodular, but they are rather technical and a detailed analysis in that direction is outside the scope of the present paper. For further details and applications of unimodular matrices the reader can refer to, e.g., Schrijver (2003).

8 Discussion

We can ask a skeptical general question: given the wealth of combinatorial theory to find orthogonal blocks what benefit does the circuit method have? An immediate answer is that it provides, in appropriate cases, the choice of a large, even very large, variety of valid randomizations schemes and under special conditions *all* valid randomizations.

Weighing designs give some intuition. Historically there are two types. Weighing a set of objects on a single pan weighing machine is very similar to the choice experiments. A chemical balance experiment has two pans and compares sets of objects. In the chemical balance the observation itself is a difference, that is an empirical contrast, whereas in the single pan case we have to reparametrized creating X_1 to obtain contrasts, as in the A/B experiment. Informally, we could say the contrast matrix X_1 represents a two-pan experiment embedded in a one pan experiment.

It is important to emphasize that the nature of the lattice of circuits in a particular problem depends on the structure of the X_1 matrix. Cost considerations and optimality of the experiment may point toward particular randomization schemes. In some cases choice of X_1 may mean there is no randomization other that full randomization (over units) of the whole experiment. Conversely, the need to randomize because of perceived sources of bias will restrict the form of X_1 as in simple A/B testing.

The blocks of a randomization scheme as defined here generalize the idea of a randomized blocked experiment and there is no requirement for equal block size,

unless imposed. Stratified sampling is covered if the contrast of interest are within strata. Valid randomizations form a lattice under refinement which we suggest is natural generalization of nested randomization. A single non-decomposable binary vector orthogonal to the X_1 matrix is a minimal element. A non-decomposable valid randomization corresponds to partition of $\mathcal{N} = \{1, 2, \ldots, n\}$. There may be more than one non-decomposable valid scheme, as we saw in the 2^3 example in Sect. 4.1 and in Example 7.

Also relevant is randomization cost. It may be that a cost function which is related to the structure of the randomization and which is order preserving with respect to the refinement in the lattice could lead to useful strategies in cases where, as we have seen, the choice of valid randomizations is very large. That is, we have in the background the idea that more refined randomization is cheaper. There is a considerable literature on sequential randomization with a model, in the A/B case, that subjects (e.g., patients) are awarded treatments A or B on the equivalent of a toss of a fair coin (there is a considerable work on biased coin design which we do not cover). This is an example where the method in this paper should be a cheaper procedure administratively than randomizing over a fixed population in order to conduct a more complex randomized block experiment. Note that in the 2^2 experiment of Example 1 with two blocks of size 2, each block only supplies some of the information. The same for the 4 blocks of size 2 in the 2^3 experiment, whereas for the two $\frac{1}{2}$ fractions of size 4 the parameters can be estimated from each block. In the 2-out-of-4 choice experiments we compare similarly attributes (1, 2) v. (3, 4), (1, 3) v. (2, 4) and (1, 4) v. (2, 3). The two-pan metaphor is useful. The extension to the k-out-of-$2k$ example is straightforward and the blocks arise from all ways of splitting $2k$ objects into disjoint set of size k. It is likely, in our view, that sequential and adaptive randomization will be increasingly important as costs are traded with effectiveness. Their impressive use in CoViD-19 vaccination trials (e.g., Thorlund et al. 2020; Knoll and Wonodi, 2021) is likely to have a lasting impact.

The paper could have been written concentrating on the link to matroid theory, indeed the term *circuit* is from matroid theory and the circuits presented here form a *linear circuit*, in the matroid sense. Another mathematical feature is that each block of randomization scheme defined here has an associated permutation group and the full randomization scheme generates a subgroup of the full permutation group S_n. All possible schemes for a particular example may lead to a complex lattice of subgroups under set partition refinement. The relation between matroids and permutation groups has been studied in Cameron and Fon-Der-Flaass (1995).

Declarations

Conflict of interest The authors declare no conflicts of interest associated with this manuscript.

References

4ti2 team. (2018). 4ti2—A software package for algebraic, geometric and combinatorial problems on linear spaces. https://4ti2.github.io.

Bailey, R. A., Rowley, C. A. (1987). Valid randomization. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences, 410*(1838), 105–124. https://doi.org/10.1098/rspa.1987.0030.

Box, G. E., Hunter, J. S., Hunter, W. J. (1978). *Statistics for experimenters*. New York: John Wiley and Sons.

Cameron, P. J., Fon-Der-Flaass, D. (1995). Bases for permutation groups and matroids. *European Journal of Combinatorics, 16*(6), 537–544. https://doi.org/10.1016/0195-6698(95)90035-7.

Cox, D. R. (2009). Randomization in the design of experiments. *International Statistical Review, 77*(3), 415–429. https://doi.org/10.1111/j.1751-5823.2009.00084.x.

Das, M., Jain, R. (1970). On component analysis of factorial and fractional factorial experiments. *Biometrics, 26*(4), 823–833.

Dean, A., Morris, M., Stufken, J., Bingham, D. (2015). *Handbook of design and analysis of experiments*. Boca Raton, FL: Chapman & Hall CRC Press.

Diaconis, P., Sturmfels, B. (1998). Algebraic algorithms for sampling from conditional distributions. *The Annals of Statistics, 26*(1), 363–397. https://doi.org/10.1214/aos/1030563990.

Drovandi, C. C., Holmes, C. C., McGree, J. M., Mengersen, K., Richardson, S., Ryan, E. G. (2017). Principles of experimental design for big data analysis. *Statistical Science, 32*(3), 385–404. https://doi.org/10.1214/16-STS604.

Fontana, R., Rapallo, F., Wynn, H. P. (2022). Circuits for robust designs. *Statistical Papers, 63*(5), 1537–1560. https://doi.org/10.1007/s00362-021-01285-6.

Grayson, D. R., Stillman, M. E. (2019). Macaulay2, a software system for research in algebraic geometry. Available at http://www.math.uiuc.edu/Macaulay2/.

Guzowski, L., Tatara, E., Milostan, C. (2014). *Scoping study using randomized controlled trials to optimize small buildings' and small portfolios'(SBSP) energy efficiency programs*. Technical Report, ANL/DIS-14/8, Argonne National Lab.(ANL), Argonne, IL (United States).

Kahle, D., O'Neill, C., Sommars, J. (2020). A computer algebra system for R: Macaulay2 and the m2r package. *Journal of Statistical Software, 93*(9), 1–31. https://doi.org/10.18637/jss.v093.i09.

Kempthorne, O. (1955). The randomization theory of experimental inference. *Journal of the American Statistical Association, 50*(271), 946–967. https://doi.org/10.2307/2281178.

Knoll, M. D., Wonodi, C. (2021). Oxford-AstraZeneca COVID-19 vaccine efficacy. *The Lancet, 397*(10269), 72–74. https://doi.org/10.1016/S2589-7500(20)30086-8.

Kohavi, R., Longbotham, R. (2017). Online controlled experiments and A/B testing. In C. Sammut, G. I. Webb (Eds.), *Encyclopedia of machine learning and data mining* (pp. 922–929). Springer. https://doi.org/10.1007/978-1-4899-7687-1_891.

Pesce, E., Riccomagno, E., Wynn, H. P. (2019). Experimental design issues in big data: The question of bias. In F. Greselin, L. Deldossi, L. Bagnato, M. Vichi (Eds.), *Statistical learning of complex data* (pp. 193–201). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-21140-0_20.

Pesce, E., Porro, F., Riccomagno, E. (2022). Large datasets, bias and model-oriented optimal design of experiments. *Quality and Reliability Engineering International*. https://doi.org/10.1002/qre.3165

Schrijver, A. (2003). *Combinatorial optimization. Polyhedra and efficiency, algorithms and combinatorics*. Berlin: Springer.

Scott, A. J., Smith, T. M. F. (1975). Minimax designs for sample surveys. *Biometrika, 62*(2), 353–357. https://doi.org/10.2307/2335372.

Simões Pereira, J. (1975). On matroids on edge sets of graphs with connected subgraphs as circuits II. *Discrete Mathematics, 12*, 55–78. https://doi.org/10.1016/0012-365X(75)90095-3.

Stenger, H. (1979). A minimax approach to randomization and estimation in survey sampling. *The Annals of Statistics, 7*(2), 395–399. https://doi.org/10.1214/aos/1176344622.

Stigler, S. M. (1969). The use of random allocation for the control of selection bias. *Biometrika, 56*(3), 553–560. https://doi.org/10.2307/2334663.

Sturmfels, B. (1996). *Gröbner bases and convex polytopes. University lecture series* (Vol. 8). Providence, RI: American Mathematical Society.

Thorlund, K., Dron, L., Park, J., Hsu, G., Forrest, J. I., Mills, E. J. (2020). A real-time dashboard of clinical trials for COVID-19. *The Lancet Digital Health, 2*(6), E286–E287. https://doi.org/10.1016/S2589-7500(20)30086-8.

Wynn, H. P. (1977). Minimax purposive survey sampling design. *Journal of the American Statistical Association, 72*(359), 655–657. https://doi.org/10.2307/2286234.