Automatic generation of regular factorial designs using the R PLANOR library

Hervé Monod, André Kobilinsky, Annie Bouvier

INRA, Unité MIA-Jouy, France

Conférence AGROSTAT

29 February 2012

Plan

Plan

Plan

Plan

Plan

Short INRA-Versailles story

Thanks to the impulsion of André Kobilinsky :

- $>$ 30 years of active research on the design of experiments at INRA-Versailles
- including many *Agro-food applications :*
  - process optimisation, food testing, ...
- $\Rightarrow$ regular factorial designs and extensions

$+$ *Agriculture and computer experiment applications*

Regular factorial design

The roots

- Yates (1933, 1937), Fisher (1942), *Rothamsted Experimental Station*, Finney (1947) : factorial designs, confounding
- ...
- Patterson (1976), Bailey (1977), Franklin (1977) : design key
- ...

A revival ?

- *Biometrics, 2011* : Split-Plot Designs for Robotic Serial Dilution Assays
- *Genetics, 2010* : Statistical design and analysis of RNA sequencing data
- *Biometrika, 2011* : Nested orthogonal array-based Latin hypercube designs
- *2011 : Annals of Statistics, JASA, JSPI, Technometrics, etc*

Factorial design : today

Software offer such as

- SAS : proc factex, + Kuhfeld and Tobias (2005) *Technometrics*
- R library FrF2 (2010)

does not satisfy simultaneously

- varied number of levels between factors
- asymetric models with respect to the factors
- hierarchical constraints between factors
- multistratum designs (*e.g.* split-plot)
- + . . . freely available

R PLANOR library

The PLANOR project

- 1994, Kobilinsky. Automatic generation of asymmetrical regular designs.
- 1995, 2005, Kobilinsky. PLANOR : program for the automatic generation of regular experimental designs.
- 2011, Kobilinsky, Bouvier, Monod. PLANOR : an R library for the automatic generation of regular fractional factorial designs

R PLANOR library

The PLANOR project

- 1994, Kobilinsky. Automatic generation of asymmetrical regular designs.
- 1995, 2005, Kobilinsky. PLANOR : program for the automatic generation of regular experimental designs.
- 2011, Kobilinsky, Bouvier, Monod. PLANOR : an R library for the automatic generation of regular fractional factorial designs

Objectives of the R version :

- to enlarge the users' community (statisticians, experimenters)

Plan

Example 1 : 4 factors at 2 levels in $N = 8$ units

$\Rightarrow$ select 8 among $2^4 = 16$ treatments

Unit	A	B	C	$D = ABC$
1	0	0	0	0
2	0	0	1	1
3	0	1	0	1
4	0	1	1	0
5	1	0	0	1
6	1	0	1	0
7	1	1	0	0
8	1	1	1	1

Consequences : model matrix

$$X = \begin{array}{c}
\begin{array}{ccccccccccc} 1 & A & B & C & D & AB & AC & AD & BC & BD & CD \end{array} \\
\left(\begin{array}{ccccccccccc}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 \\
1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 \\
1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\
1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 \\
1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 \\
1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1
\end{array}\right)
\end{array}$$

Effets factoriels *confounded (aliased)* 2 by 2

Consequences : aliasing

$$\begin{aligned} 1 &= ABCD \\ A &= BCD \\ B &= ACD \\ C &= ABD \\ D &= ABC \\ AB &= CD \\ AC &= BD \\ AD &= BC \end{aligned}$$

null interactions of order $\geq 3 \Rightarrow$ main effects are estimable

Consequences : modelling

In the model :

$$\begin{aligned} Y &= 1 + A + B + C + D \\ &\quad + A.B + A.C + A.D + B.C + B.D + C.D \\ &\quad + \varepsilon \end{aligned}$$

All main effects are estimable

Key matrix

Unit	*A*	*B*	*C*	$D = ABC$
000	0	0	0	0
001	0	0	1	1
010	0	1	0	1
011	0	1	1	0
100	1	0	0	1
101	1	0	1	0
110	1	1	0	0
111	1	1	1	1

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Plan

Search for a design $\Longrightarrow$ inverse problem

Starting point :

1. list of factors and their levels
2. model(s) and list(s) of terms to estimate
3. number of experimental units

Search for a design $\Longrightarrow$ inverse problem

Starting point :

1. list of factors and their levels
2. model(s) and list(s) of terms to estimate
3. number of experimental units

End point

1. experimental design
2. randomization

Search for a design $\Longrightarrow$ inverse problem

Starting point :

1. list of factors and their levels
2. model(s) and list(s) of terms to estimate
3. number of experimental units

PLANOR $\Longrightarrow$ search for a key matrix

End point

1. experimental design
2. randomization

Plan

Sensory experiment

Facteurs :

- 6 tasters (D)
- 6 periods (P) in 2 series (S)
- 2 recipies (A)
- 3 presentations (B)

Modèle :

- $\text{Taster} + \text{Serie}/\text{Period} + A + B + A.B$
- Estimate A, B, $A.B$

Method in practice : example 2

```
> library("planor",lib="/home/hmonod/R/library")

> deguste.key <- planor.designkey(
   factors = c("D","P","S","A","B"),
   nlevels = c(6,6,2,2,3),
   hiera = ~S/P,
   model = ~ D + S*P + A*B, estimate = ~A*B,
   nunits = 36,
   base = ~D+P)
```

Information while running

```
Determination of ineligible factorial terms
Determination of ineligible pseudofactorial terms
Independent searches for prime(s) :  2 3
Key-matrix search for prime p = 2
There are  2  predefined columns
First visit to column  3
First visit to column  4
The search is closed: max.sol =  1 solution(s) found
Key-matrix search for prime p = 3
There are  2  predefined columns
First visit to column  3
The search is closed: max.sol =  1 solution(s) found
```

Intermediate results : key matrixes

```
> print(deguste.key)
--- Solution  1  for prime  2  ---

    D_1 P_1 S A
D_1   1   0 0 1
P_1   0   1 1 1

--- Solution  1  for prime  3  ---

    D_2 P_2 B
D_2   1   0 1
P_2   0   1 1
```

Final result : randomized design I

```
> deguste.plan <- planor.design(deguste.key,
                               randomize=~D + S/P)

   D P S A B
1  1 1 1 1 3
2  1 2 1 1 2
3  1 3 1 1 1
...
```

1c	1b	1a	2c	2a	2b
1a	1c	1b	2a	2b	2c
2c	2b	2a	1c	1a	1b
2a	2c	2b	1a	1b	1c
2b	2a	2c	1b	1c	1a
1b	1a	1c	2b	2c	2a

Last (and larger) example 3 : computer experiment

Sensitivity analysis of a model in animal epidemiology

(Courcoul, Monod *et al.*, 2011, J. Theor. Biol.)

- 12 factors at 4 levels
- 7 factors at 2 levels
- Model : main effects + two-factor nteractions

$\Longrightarrow 4^{12} \times 2^{7} = 2^{31}$ possible combinations

Solution can be found with $2^{12} = 4096$ simulations

Plan

Discussion

Main characteristics

- an R library, documented and available from MIA-Jouy website
- generates orthogonal *and* regular designs only, but
  - with any numbers of levels
  - possibly with hierarchical hierarchies
  - several levels of variability taken into account (split-plot)
- not optimised for computing time but
  - can manage large designs

Discussion

Main characteristics

- an R library, documented and available from MIA-Jouy website
- generates orthogonal *and* regular designs only, but
  - with any numbers of levels
  - possibly with hierarchical hierarchies
  - several levels of variability taken into account (split-plot)
- not optimised for computing time but
  - can manage large designs

Perspectives

- improve overalll performances
- integrate other algorithmes than backtrack
- refine selection criteria
- adapt to sampling designs in high dimensions

Discussion

Main characteristics

- an R library, documented and available from MIA-Jouy website
- generates orthogonal *and* regular designs only, but
  - with any numbers of levels
  - possibly with hierarchical hierarchies
  - several levels of variability taken into account (split-plot)
- not optimised for computing time but
  - can manage large designs

Perspectives

- improve overalll performances
- integrate other algorithmes than backtrack
- refine selection criteria
- adapt to sampling designs in high dimensions

Testers welcome !

Collaborations

- André Kobilinsky (INRA, MIA-Jouy)

- $+$ Rosemary Bailey (Queen Mary, University of London)