

Plans factoriels réguliers

Hervé Monod



Pec'num²⁰¹⁵

École Centrale de Lyon, 19 mai 2015

Introduction

- Bref historique
- Périmètre de l'exposé

Plans pour facteurs qualitatifs

- Plans en blocs orthogonaux
- Plans multi-factoriels

Génération automatique de plans factoriels réguliers

- Principe
- Cadre théorique
- Application à la recherche de plans en blocs
- Application à la recherche de plans de grandes tailles

Plans pour facteurs quantitatifs

- Principe
- Construction

Discussion

Introduction

- Bref historique

- Périmètre de l'exposé

Plans pour facteurs qualitatifs

- Plans en blocs orthogonaux

- Plans multi-factoriels

Génération automatique de plans factoriels réguliers

- Principe

- Cadre théorique

- Application à la recherche de plans en blocs

- Application à la recherche de plans de grandes tailles

Plans pour facteurs quantitatifs

- Principe

- Construction

Discussion

des prémisses

- ▶ Young (1771), Cretté de Palluel (1788): soulignent l'intérêt des expériences comparatives; élaborent des protocoles

19e siècle: l'essor de la méthode expérimentale

- ▶ Johnson *Experimental Agriculture* (1849)
- ▶ Claude Bernard *Introd. à l'étude de la médecine exp.* (1865)

20e siècle: la théorie statistique des plans d'expériences

1. **agriculture, industrie** Fisher *Statistical methods for research workers* (1925); *The design of experiments* (1935)
2. **médecine** Austin Bradford Hill (1946): 1er essai clinique randomisé (efficacité de la streptomycine)
3. **industrie, chimie** G.E.P. Box, Taguchi, ...

L'expérimentation en agriculture



Gérer la variabilité expérimentale: les 3 R de Fisher

- ▶ **R**épétitions
- ▶ **B**locs
- ▶ **R**andomisation

Étudier plusieurs facteurs simultanément

- ▶ plus efficace
- ▶ permet de détecter des interactions

Optimiser (Kiefer, 1960)

- ▶ la précision des paramètres, des prédictions
- ▶ sous différentes conditions (robustesse)

Introduction

Bref historique

Périmètre de l'exposé

Plans pour facteurs qualitatifs

Plans en blocs orthogonaux

Plans multi-factoriels

Génération automatique de plans factoriels réguliers

Principe

Cadre théorique

Application à la recherche de plans en blocs

Application à la recherche de plans de grandes tailles

Plans pour facteurs quantitatifs

Principe

Construction

Discussion

Dans cet exposé **Plans factoriels réguliers**

⇔ plans orthogonaux, basés sur des méthodes algébriques de construction

Objectifs

- ▶ présenter les grands principes des plans factoriels discrets
- ▶ montrer leur généralité
- ▶ faire le pont avec des plans space-filling
- ▶ décrire un algorithme **générique** pour construire et randomiser ces plans
- ▶ et son implémentation dans le package R PLANOR

- ▶ $A, B, \dots$ ou $F_1, F_2, \dots$: facteurs d'entrée
- ▶ n : nombre de facteurs d'entrée
- ▶ m_F ou m_i : nombre de modalités du facteur F ou F_i
- ▶ les modalités d'un facteur à m_i sont notées $0, 1, \dots, m_i$
- ▶ N : nombre total d'unités expérimentales (ou *taille* du plan)
- ▶ U : facteur(s) associé(s) aux unités expérimentales

Introduction

- Bref historique

- Périmètre de l'exposé

Plans pour facteurs qualitatifs

- Plans en blocs orthogonaux

- Plans multi-factoriels

Génération automatique de plans factoriels réguliers

- Principe

- Cadre théorique

- Application à la recherche de plans en blocs

- Application à la recherche de plans de grandes tailles

Plans pour facteurs quantitatifs

- Principe

- Construction

Discussion

Objectif

Comparer les modalités d'un facteur traitement en contrôlant les effets d'un facteur bloc

- ▶ n traitements **répétés** r fois
- ▶ Chaque **bloc** contient chaque traitement 1 fois
- ▶ ordre aléatoire des traitements dans les blocs
- ▶ Analyse selon le modèle: $y_{ij} = \beta_j + \tau_i + \varepsilon_{ij}$; $\text{Var}(\varepsilon_{ij}) = \sigma^2$

Exemple agricole:

6 variétés 0, ..., 5

4 blocs

4	2	0	5	3	1
1	4	3	2	5	0
3	4	2	1	5	0
1	3	5	2	4	0

Plan en blocs complets randomisés: construction

1. Construction du plan initial:

- ▶ $n = 2$ facteurs B et V ; $m_B = 4$, $m_V = 6$
- ▶ $N = m_B \times m_V = 24$ unités
- ▶ Chaque paire (b, v) présente une fois

U	B	V
1	0	0
2	0	1
3	0	2
$\vdots$	$\vdots$	
23	3	4
24	3	5

2. Randomisation: $\sim$ bloc / parcelle

Bloc 0	0	1	2	3	4	5
Bloc 1	0	1	2	3	4	5
Bloc 2	0	1	2	3	4	5
Bloc 3	0	1	2	3	4	5

$\Rightarrow$

4	2	0	5	3	1
1	4	3	2	5	0
3	4	2	1	5	0
1	3	5	2	4	0

Plan en blocs complets randomisés: analyse statistique

Analyse

- ▶ Modèle: $y_{ij} = \beta_j + \tau_i + \varepsilon_{ij}$; $\text{Var}(\varepsilon_{ij}) = \sigma^2$
- ▶ Analyse de la variance
- ▶ Estimation des paramètres et comparaison des $\hat{\tau}_i$

Tableau d'analyse de la variance

Terme factoriel	ddl	ddl (ex)	SC	CM	F	P
Bloc	$m_B - 1$	3	SC_B	CM_B	F_B	P_B
Variété	$m_V - 1$	5	SC_V	CM_V	F_V	P_V
Résiduelle	$(m_B - 1)(m_V - 1)$	15	SCR	CMR		

Remarque: $(m_B - 1)(m_V - 1) = N - (m_B - 1) - (m_V - 1) - 1$

Propriétés statistiques

- ▶ tous les effets principaux sont estimables; orthogonalité entre estimateurs
- ▶ plan optimal: variance minimale de la moyenne des $\hat{\tau}_i - \hat{\tau}_j$ parmi les plans de taille N
- ▶ interaction blocs x variétés confondue avec la résiduelle
- ▶ validité du modèle apportée par la randomisation

Exemple 2: plan en carré latin

Objectif

Comparer les modalités d'un facteur traitement en contrôlant les effets de deux facteurs blocs croisés

- Facteurs: traitement (T , $m_T = m$), juge ($m_L = m$), période ($m_C = m$)
- $N = m^2$. chaque ligne et chaque colonne contiennent chaque traitement 1 fois
- Analyse selon le modèle: $Y \sim \text{juge} + \text{période} + \text{produit} + \varepsilon$

Exemple: expérience de dégustation

5 produits 0, ..., 4

5 juges $\times$ 5 périodes

		Période				
Juge		0	1	2	4	3
		3	4	0	2	1
		2	3	4	1	0
		4	0	1	3	2
		1	2	3	0	4

Plan en carré latin: construction

1. Construction du plan initial:

- ▶ $n = 3$ facteurs J, P, T ; $m_J = m_P = m_T = 5$
- ▶ $N = m^2 = 25$ unités
- ▶ $T = J + P \pmod{5}$
- ▶ Chaque triplet (j, p, t) présent une ou zéro fois
- ▶ Chaque paire (j, p) , (j, t) ou (p, t) présente une fois

U	J	P	T
1	0	0	0
2	0	1	1
3	0	2	2
$\vdots$	$\vdots$		
24	4	3	2
25	4	4	3

2. Randomisation: $\sim$ ligne $\times$ colonne

		Période						
Juge		0	1	2	3	4	$\Rightarrow$	
	1	1	2	3	4	0		
	2	2	3	4	0	1		
	3	3	4	0	1	2		
	4	4	0	1	2	3		
		Période						
Juge		0	1	2	4	3		
	3	3	4	0	2	1		
	2	2	3	4	1	0		
	4	4	0	1	3	2		
	1	1	2	3	0	4		

Plan en carré latin: analyse statistique

Analyse

- ▶ Modèle: $y_{ijt} = \lambda_i + \gamma_j + \tau_t + \varepsilon_{ijt}$; $\text{Var}(\varepsilon_{ijt}) = \sigma^2$
- ▶ Analyse de la variance
- ▶ Estimation des paramètres et comparaison des $\hat{\tau}_t$

Tableau d'analyse de la variance

Terme factoriel	ddl	ddl (ex)	SC	CM	F	P
Ligne	$m - 1$	4	SC_L	CM_L	F_L	P_L
Colonne	$m - 1$	4	SC_C	CM_C	F_C	P_C
Traitement	$m - 1$	4	SC_T	CM_T	F_T	P_T
Résiduelle	$(m - 1)(m - 2)$	12	SCR	CMR		

Remarque: $(m - 1)(m - 2) = N - 3(m - 1) - 1$

Propriétés statistiques

- ▶ tous les effets principaux sont estimables; orthogonalité entre estimateurs
- ▶ plan optimal
- ▶ interactions partiellement confondues avec la résiduelle, les effets principaux
- ▶ validité du modèle apportée par la randomisation

Exemple 3: plan de type Sudoku

- Facteurs: Ligne, Colonne, Bloc, Traitement
- $N = m^2$; chaque ligne, chaque colonne, chaque bloc contient chaque traitement 1 fois

1	2	3	4	5	6	7	8	9
4	5	6	7	8	9	1	2	3
7	8	9	1	2	3	4	5	6
2	3	1	5	6	4	8	9	7
5	6	4	8	9	7	2	3	1
8	9	7	2	3	1	5	6	4
3	1	2	6	4	5	9	7	8
6	4	5	9	7	8	3	1	2
9	7	8	3	1	2	6	4	5

Exemple 3: plan de type Sudoku

- ▶ $T_1 = L_1 + C_2 \pmod{3}$; $T_2 = L_2 + C_1 \pmod{3}$
- ▶ $T = T_1 + 3T_2 + 1$
- ▶ $B = L_1 + 3C_1 + 1$

00	10	20	01	11	21	02	12	22
01	11	21	02	12	22	00	10	20
02	12	22	00	10	20	01	11	21
10	20	00	11	21	01	12	22	02
11	21	01	12	22	02	10	20	00
12	22	02	10	20	00	11	21	01
20	00	10	21	01	11	22	02	12
21	01	11	22	02	12	20	00	10
22	02	12	20	00	10	21	01	11

Problème: pas de randomisation valide

Introduction

- Bref historique

- Périmètre de l'exposé

Plans pour facteurs qualitatifs

- Plans en blocs orthogonaux

- Plans multi-factoriels

Génération automatique de plans factoriels réguliers

- Principe

- Cadre théorique

- Application à la recherche de plans en blocs

- Application à la recherche de plans de grandes tailles

Plans pour facteurs quantitatifs

- Principe

- Construction

Discussion

Plans factoriels fractionnaires

Plans factoriels fractionnaires

- ▶ n facteurs à $m_1, \dots, m_n$ niveaux
- ▶ budget de N simulations
- ▶ $\prod m_i \gg N$

⇒ sélectionner N combinaisons d'entrée parmi les $\prod m_i$ possibilités

Exemple: carré latin

- ▶ $n = 3, m = 5 \Rightarrow |T| = 125$
- ▶ $N = 25$

Exemple 4: 4 facteurs à 2 niveaux en $N = 8$ unités

En entrée:

- Facteurs: A, B, C, D (2)
- $N = 8 = 2^3$ unités

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

Construction

- plan initial complet pour A, B, C et $D = A + B + C \pmod{2}$
- randomisation totale

⇒ **Tableau orthogonal de force 3:**

le plan est complet pour tout sous-ensemble de 3 colonnes

Analyse

- Modèle:
$$Y \sim A + B + C + D + A.B + A.C + A.D + B.C + B.D + C.D + \varepsilon$$

Conséquences: matrice X du modèle de régression $Y = X\theta + \varepsilon$:

$$X = \begin{pmatrix} 1 & A & B & C & D & AB & AC & AD & BC & BD & CD \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Effets factoriels *confondus* (*aliased*) 2 par 2

Propriétés statistiques

Relations d'alias entre effets factoriels

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

interactions d'ordre ≥ 3 nulles $\Rightarrow$ effets principaux estimables

Conséquences: modélisation

Dans le modèle:

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

Tous les effets principaux sont estimables.

Plan fractionnaire de grande dimension

Contexte: analyse de sensibilité en épidémiologie animale

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

En entrée:

- ▶ 19 facteurs:
 - ▶ $X_1, \dots, X_{12}$: 4 niveaux
 - ▶ $X_{13}, \dots, X_{19}$: 2 niveaux
- ▶ budget: $N \leq 5000$ simulations (pour 2^{31} combinaisons possibles)
- ▶ Modèle: effets principaux et interactions entre 2 facteurs

Construction $\Rightarrow$

- ▶ plan initial complet pour $X_1, \dots, X_6$ + 19 relations à trouver
- ▶ randomisation inutile (cf. simulations)

Solution:

```
> frac.plan@design
      A B C D E F G H I J K L M N O P Q R S
1      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2      1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1
.      .
.      .
.      .
4096 4 4 4 4 1 4 2 2 2 2 4 3 2 1 1 2 1 1 1
>
```

- ▶ **Tableau orthogonal** de force 4 ($\equiv$ résolution 5)
Propriété combinatoire: plan complet pour tout sous-ensemble de 4 facteurs
- ▶ *Propriété statistique:* effets principaux confondus avec des interactions entre au moins 3 facteurs; interactions entre 2 facteurs non confondues
on peut estimer tous les effets dans un modèle d'ordre 2

Résolution d'une fraction factorielle

Tableau orthogonal (définition)

Un tableau orthogonal de force f à N lignes, n colonnes et m_i modalités par colonne i est tel que, pour tout sous-ensemble de f facteurs, toutes les combinaisons de modalités sont présentes et équilibrées.

Résolution (définition)

Un plan factoriel fractionnaire régulier est de résolution R si les effets factoriels confondus avec la moyenne générale sont des interactions entre au moins R facteurs..

Propriétés

Si P est un plan factoriel fractionnaire régulier de résolution R , alors P est un tableau orthogonal de force $f = R - 1$.

Propriétés

Dans un plan factoriel de résolution R

- ▶ les effets principaux sont confondus avec des interactions entre au moins $R - 1$ facteurs
- ▶ les interactions entre k facteurs ($k \leq R/2$) sont confondues avec des interactions entre au moins $R - k$ facteurs.

Corollaire

- ▶ Avec un plan factoriel de résolution R impair, on peut estimer tous les termes d'un modèle d'ordre $(R - 1)/2$
- ▶ Avec un plan factoriel de résolution R pair, on peut estimer tous les termes d'ordre $< R/2$ d'un modèle d'ordre $R/2$

Introduction

- Bref historique

- Périmètre de l'exposé

Plans pour facteurs qualitatifs

- Plans en blocs orthogonaux

- Plans multi-factoriels

Génération automatique de plans factoriels réguliers

- Principe

- Cadre théorique

- Application à la recherche de plans en blocs

- Application à la recherche de plans de grandes tailles

Plans pour facteurs quantitatifs

- Principe

- Construction

Discussion

Dans ce qui précède:

construction $\leftarrow$ propriétés combinatoires $\leftarrow$ modèle statistique adapté

En pratique, pour des problèmes plus complexes, il faut inverser:

modèle souhaité $\leftarrow$ propriétés combinatoires requises $\leftarrow$ construction

Introduction

- Bref historique

- Périmètre de l'exposé

Plans pour facteurs qualitatifs

- Plans en blocs orthogonaux

- Plans multi-factoriels

Génération automatique de plans factoriels réguliers

- Principe

- Cadre théorique

- Application à la recherche de plans en blocs

- Application à la recherche de plans de grandes tailles

Plans pour facteurs quantitatifs

- Principe

- Construction

Discussion

Example 4: plan fractionnaire simple

<i>Unit</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D = ABC</i>
$u_1 u_2 u_3$	t_1	t_2	t_3	t_4
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

Construction

$$\begin{pmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \textcolor{brown}{1} & \textcolor{brown}{1} & \textcolor{brown}{1} \end{pmatrix}}_{\Phi^t} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$$

Φ = est appelée matrice clé

[On considère n facteurs $F_1, \dots, F_n$ à p modalités et p^r unités:]

Traitements et unités $\implies$ *structure de groupe abélien*

- ▶ $T = (C_p)^n$; traitements: $\mathbf{t} = (t_1, \dots, t_n)$, avec $t_i \in C_p$
- ▶ $U = (C_p)^r$; unités: $\mathbf{u} = (u_1, \dots, u_s)$, avec $u_i \in C_p$

Plan régulier $\implies$ homomorphisme

$$\begin{aligned} d &: U \longrightarrow T \\ \mathbf{u} &\longmapsto d(\mathbf{u}) = \Phi^t \mathbf{u} \end{aligned}$$

with $d(\mathbf{u} + \mathbf{u}') = d(\mathbf{u}) + d(\mathbf{u}')$

Effects factoriels $\Rightarrow$ *structure de groupe abélien dual*

- ▶ $T^* = (C_p)^n$; caractères: $\mathbf{a} = (a_1, \dots, a_n)$, avec $a_i \in C_p$
- ▶ $U^* = (C_p)^r$; caractères: $\mathbf{v} = (v_1, \dots, v_s)$, avec $v_i \in C_p$

- **Chaque caractère $\mathbf{a}$ est associé à l'effet factoriel défini par:**

$$e(\mathbf{a}) = \frac{1}{\|T\|} \sum_{\mathbf{t}} \eta_p^{-\sum_i a_i \cdot t_i} \cdot \tau_{\mathbf{t}}$$

- $e(\mathbf{a})$ appartient à l'interaction entre les facteurs F_i tels que $a_i \neq 0$

Plan factoriel régulier

Une définition et une proposition clés:

Definition (plan factoriel régulier)

La fonction d de U dans T est un plan factoriel **régulier** si

$$d(\mathbf{u}) = \Phi^t \mathbf{u} + \mathbf{t}_0 \text{ avec } \Phi^t(\mathbf{u} + \mathbf{u}') = \Phi^t \mathbf{u} + \Phi^t \mathbf{u}'.$$

La matrice Φ est appelée la **matrice clé** du plan.

Proposition (Confusion d'effets)

Les **effets factoriels** associés aux caractères $\mathbf{a}$ et $\mathbf{a}'$ sont **confondus** si et seulement si $\Phi(\mathbf{a} - \mathbf{a}') = 0$

⇒ Toutes les propriétés de confusion d'effets sont déterminées par $\ker(\Phi)$.

Example 4: effets confondus

Les effets factoriels $A^a B^b C^c D^d$ and $A^{a'} B^{b'} C^{c'} D^{d'}$ sont confondus ssi

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}}_{\text{Matrice clé } \Phi} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}}_{\text{Matrice clé } \Phi} \begin{pmatrix} a' \\ b' \\ c' \\ d' \end{pmatrix}$$

Exemple: AB est confondu avec CD puisque

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

La difficulté: inverser le problème

Définir les facteurs et le modèle en entrée:

- ▶ Facteurs: A, B, C, D (2 levels)
- ▶ Modèle: $\sim A + B + C + D + A.B + \dots + C.D$
- ▶ $N = 8$

Obtenir le plan en sortie:

<i>Unit</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
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

Pour cela: trouver une matrice clé adéquate Φ automatiquement

Packages R pour construire des plans factoriels réguliers: FrF2, planor

Caractéristiques de planor:

- ▶ package R, développé à MIA-Jouy, déposé sur le CRAN
- ▶ construit et randomise des plans factoriels réguliers
- ▶ en particulier des plans factoriels **fractionnaires**
- ▶ pour expériences réelles ou numériques

Ensemble non éligible

Dans les définitions ci-dessous, un terme factoriel est identifié à un sous-ensemble de facteurs. Les propositions sont vraies sous des conditions techniques généralement vérifiées en pratique; voir Kobilinsky, Monod et Bailey (2015) pour plus de détails.

Définitions

Pour un ensemble de facteurs $F_1 \dots F_n$ et un modèle M composé de termes factoriels donnés sur ces facteurs, on définit les sous-ensembles suivants de termes factoriels:

- ▶ $\mathcal{M}$ (Model): termes du modèle
- ▶ $\mathcal{E}$ (Estimate): termes du modèle à estimer
- ▶ $\mathcal{I} = \{\mathbf{a} \cup \mathbf{b} : \mathbf{a} \in \mathcal{E}, \mathbf{b} \in \mathcal{M}\}$ (Ineligible): termes factoriels inéligibles

Exemple avec $p = 2$:

- ▶ Modèle: $Y\tilde{A} + B + C + D + A.B + A.C$
- ▶ $\mathcal{M} = \{1, A, B, C, D, A.B, A.C\}$
- ▶ $\mathcal{E} = \{A, B, C\}$
- ▶ $\mathcal{I}_{\mathcal{F}} = \{A, B, C, A.B, A.C, A.D, B.C, B.D, C.D, A.B.C\}$

Version combinatoire

On considère le modèle composé des termes factoriels dans $\mathcal{M}$ et le sous-ensemble $\mathcal{E}$ des termes à estimer. Tous les effets factoriels de $\mathcal{E}$ sont estimables si, pour tout terme factoriel $F_{j_1} \dots F_{j_k}$ dans $\mathcal{I}$, chaque combinaison de modalités des facteurs $F_{j_1} \dots F_{j_k}$ est présente au moins une fois.

$$\ker(\Phi) \cap \mathcal{I} = \emptyset.$$

Cas des plans factoriels réguliers

On considère le modèle $\mathcal{M}$ et un plan factoriel régulier dont la matrice clé est Φ . Soit $\mathcal{I}_C$ l'ensemble des caractères associés aux termes factoriels dans $\mathcal{I}$. Tous les effets factoriels de $\mathcal{E}$ sont estimables si et seulement si

$$\ker(\Phi) \cap \mathcal{I}_C = \emptyset.$$

Algorithme implémenté dans Planor

Principales étapes:

1. décomposition des facteurs en **pseudofacteurs** ayant un nombre premier de modalités p
2. décomposition des termes factoriels dans $\mathcal{M}$ et $\mathcal{E}$ en termes pseudofactoriels et détermination de $\mathcal{I}$
3. pour chaque premier p
 - 3.1 calcul de $\mathcal{I}(p)$ et détermination du sous-ensemble de caractères $\mathcal{I}_C(p)$
 - 3.2 recherche de Φ_p colonne par colonne (algorithme backtrack), de façon que $\ker(\Phi_p) \cap \mathcal{I}$ reste vide
4. (possiblement) comparaison des solutions
5. construction et randomisation du plan en utilisant la matrice clé Φ_p

Introduction

- Bref historique

- Périmètre de l'exposé

Plans pour facteurs qualitatifs

- Plans en blocs orthogonaux

- Plans multi-factoriels

Génération automatique de plans factoriels réguliers

- Principe

- Cadre théorique

- Application à la recherche de plans en blocs

- Application à la recherche de plans de grandes tailles

Plans pour facteurs quantitatifs

- Principe

- Construction

Discussion

Sensory experiment

Facteurs:

- ▶ 2 recipes (A) $\times$ 3 presentations (B)
- ▶ 6 tasters (D)
- ▶ 6 periods (P) in 2 series (S)

Modèle:

- ▶ Taster + Serie/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

Introduction

- Bref historique

- Périmètre de l'exposé

Plans pour facteurs qualitatifs

- Plans en blocs orthogonaux

- Plans multi-factoriels

Génération automatique de plans factoriels réguliers

- Principe

- Cadre théorique

- Application à la recherche de plans en blocs

- Application à la recherche de plans de grandes tailles

Plans pour facteurs quantitatifs

- Principe

- Construction

Discussion

Exemple 5: plan fractionnaire de grande dimension

Contexte: analyse de sensibilité en épidémiologie animale

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

En entrée:

- ▶ 19 facteurs:
 - ▶ $X_1, \dots, X_{12}$: 4 niveaux
 - ▶ $X_{13}, \dots, X_{19}$: 2 niveaux
- ▶ budget: $N \leq 5000$ simulations (pour 2^{31} combinaisons possibles)
- ▶ Modèle: effets principaux et interactions entre 2 facteurs

Construction $\Rightarrow$

- ▶ plan initial complet pour $X_1, \dots, X_6$ + 19 relations à trouver
- ▶ randomisation inutile (cf. simulations)

Application de planor (exemple 5)

Plan factoriel complet: $4^{12} \times 2^7 = 2^{31}$ combinaisons ($\approx 2.15 \cdot 10^9$)

On cherche: un plan en $N = 2^{12} = 4096$

Syntaxe **planor**:

```
> set.seed(123)
> frac.plan <- regular.design(
  factors = LETTERS[1:19],
  nlevels = c(rep(4,12),rep(2,7)),
  resolution = 5,
  nunits = 4096,
  base = ~ A + B + C + D + E + F)
[...]
```

Solution (55 secondes sur PC):

```
> frac.plan@design
      A B C D E F G H I J K L M N O P Q R S
1      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2      1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1
.      .
.      .
.      .
4096 4 4 4 4 1 4 2 2 2 2 4 3 2 1 1 2 1 1 1
>
```

⇒ **Tableau orthogonal** de force 4 ($\equiv$ résolution 5)

Propriété: plan complet pour tout sous-ensemble de 4 facteurs

Introduction

- Bref historique

- Périmètre de l'exposé

Plans pour facteurs qualitatifs

- Plans en blocs orthogonaux

- Plans multi-factoriels

Génération automatique de plans factoriels réguliers

- Principe

- Cadre théorique

- Application à la recherche de plans en blocs

- Application à la recherche de plans de grandes tailles

Plans pour facteurs quantitatifs

- Principe

- Construction

Discussion

Plans d'échantillonnage "space-filling"

- ▶ n facteurs
- ▶ domaines de définition continus: $[0, 1]$ sans perte de généralité
- ▶ budget de N simulations

⇒ échantillonner N combinaisons d'entrée dans $[0, 1]^n$

Contrainte supplémentaire par rapport au cas discret:
bien explorer les marginales

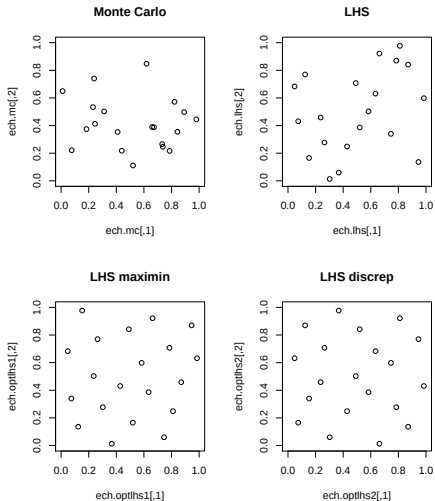
Critères classiques

- ▶ distances entre points
- ▶ discrédance
- ▶ propriétés combinatoires

Grande variété d'approches

- ▶ échantillonnage par Monte Carlo
- ▶ hypercube latin $\Rightarrow$ bonne exploration des marginales
- ▶ séquences de quasi-Monte Carlo
- ▶ plans “uniformes” $\Rightarrow$ optimisent la distance entre points

Méthodes d'échantillonnage: quelques exemples



Introduction

- Bref historique

- Périmètre de l'exposé

Plans pour facteurs qualitatifs

- Plans en blocs orthogonaux

- Plans multi-factoriels

Génération automatique de plans factoriels réguliers

- Principe

- Cadre théorique

- Application à la recherche de plans en blocs

- Application à la recherche de plans de grandes tailles

Plans pour facteurs quantitatifs

- Principe

- Construction

Discussion

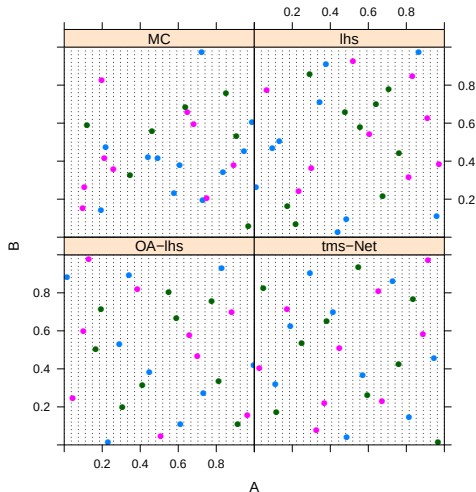
Principe de construction (cas de 3 facteurs avec $N = 27$)

- ▶ on découpe $[0, 1]$ en 27 sous-intervalles $S_i = [i/27, (i+1)/27[$, pour $i = 0, \dots, 26$
- ▶ on associe au facteur A trois pseudofacteurs A_1, A_2, A_3 de niveaux 0, 1, 2
- ▶ idem pour B et C
- ▶ on cherche une matrice clé et on construit un plan pour $A_1, \dots, C_3$
- ▶ on associe à (a_1, a_2, a_3) un tirage aléatoire dans $S_{9a_1+3a_2+a_3}$

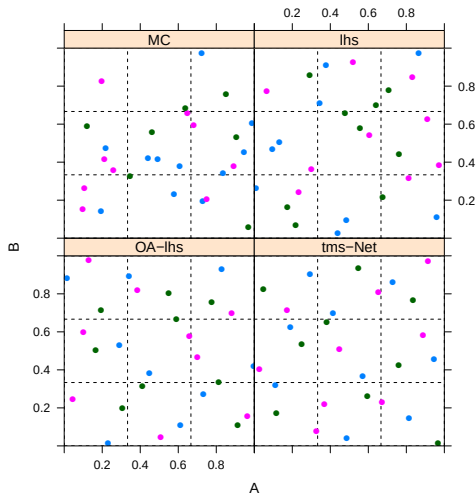
Conséquences:

- ▶ LHS $\Leftrightarrow$ plan complet pour $\{A_1, A_2, A_3\}, \{B_1, B_2, B_3\}, \{C_1, C_2, C_3\}$
- ▶ OA $\Leftrightarrow$ plan complet pour $\{A_1, B_1, C_1\}$
- ▶ $+(t, m, s)$ -net $\Leftrightarrow$ plan complet pour $\{A_1, A_2, B_1\}, \{A_1, A_2, C_1\}, \{B_1, B_2, A_1\}, \{B_1, B_2, C_1\}, \{C_1, C_2, A_1\}, \{C_1, C_2, B_1\}$

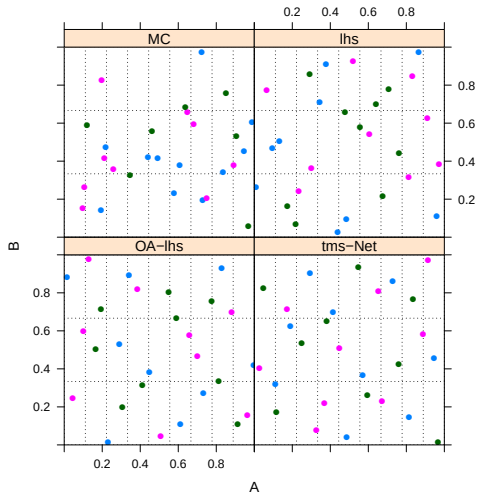
Exemple: 3 facteurs à 3 niveaux, $N = 27$



Exemple: 3 facteurs à 3 niveaux, $N = 27$



Exemple: 3 facteurs à 3 niveaux, $N = 27$



Grande cohérence entre des familles de plans d'expériences utilisées dans des contextes très différents

- ▶ noyau combinatoire et algébrique commun
- ▶ pré- et post-traitements adaptés aux différents types de plans

Échantillonnage basé sur les plans factoriels réguliers:

- ▶ génère des plans équivalents à certaines séquences de quasi-Monte Carlo: séquences de Sobol, de Niederreiter, etc.
 - ▶ mais possibilité de spécifications plus fines
 - ▶ applicable à des mélanges de facteurs qualitatifs et quantitatifs
- ▶ quelques contraintes pratiques
 - ▶ restriction à des plans orthogonaux
 - ▶ contraintes sur le nombre d'essais N
- ▶ et des avantages
 - ▶ propriétés algébriques précises des plans obtenus
 - ▶ implémentation naturelle de techniques de randomisation (brouillage)
- ▶ utilisables sous R (package planor + scripts Ecole-chercheurs Aspen)

Des logiciels existent, certains depuis longtemps

- ▶ DSIGN (≤ 1965) (*but needed the design key*)
- ▶ AKEY, FKEY in Genstat (see Payne, 2012)
- ▶ SAS: proc factex (Kuhfeld and Tobias; 2005)
- ▶ R library FrF2 (Ulrike Groemping, 2010-2012)

mais ne satisfont pas simultanément

- ▶ nombres divers de modalités des facteurs
- ▶ modèles asymétriques
- ▶ contraintes de hiérarchie entre facteurs
- ▶ plans multistrates (e.g. split-plot)
- ▶ randomisation
- ▶ + ... libre et gratuit

- ▶ Collaboration avec André Kobilinsky (†2/12/2013), Rosemary Bailey, Annie Bouvier
- ▶ Effectué en partie à l'Isaac Newton Institute of Cambridge (DAE program)

PLans factoriels réguliers

- ▶ Yates (1933, 1937), Fisher (1942): factorial designs with confounding
- ▶ Patterson (1976), Bailey (1977), Franklin (1977): **design key**
- ▶ Kobilinsky (1991), Kobilinsky et Monod (1991), Pistone and Rogantin (2008): mixed fractional factorial designs
- ▶ Kobilinsky, Monod, Bailey (2015). Automatic generation of regular factorial designs. <https://www.newton.ac.uk/files/preprints/ni15007.pdf>

Quelques recherches récentes

- ▶ *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