Codes numériques : la méthode des plans d'expériences est la pire à l'exception de toutes les autres

Jean-Marc AZAÏS

Modélisation **A**léatoire à **F**inalité **I**ndustrielle et **A**ppliquée, LSP , IMT Toulouse

La méthode des plans d'expériences est en principe une méthode en environnement incertain qui n'est pas faite pour les codes numériques. Pourtant c'est souvent la seule méthode possible

Outline

Historique

La méthode des plans d'expériences (Ronald Fisher 1920) : Comparer des variétés ou des traitements du type engrais : Le traitement .
Champs découpé en parcelles. Le problème : contrôle de la fertilité du champ. Jusque là

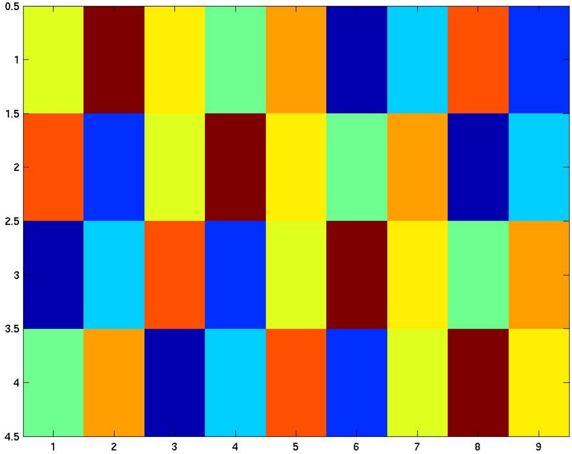
- la méthode des parcelles "témoin" :
  Le rendement est estimé par différence à la parcelle recevant le témoin la plus proche. (Wiancko, Arny et Salmon 1921).
- la méthode de la marche du cavalier (Tedin 1931). On utilise des micro-parcelles, chaque variété est observée sur plusieurs micro-parcelles qui se déduisent l'une de l'autre par le déplacement du cavalier aux échecs.

Historique

La méthode des plans d'expériences (Ronald Fisher 1920) :
Comparer des variétés ou des traitements du type engrais : Le traitement .
Champs découpé en parcelles. Le problème : contrôle de la fertilité du champ. Jusque là

- la méthode des parcelles "témoin" :
  Le rendement est estimé par différence à la parcelle recevant le témoin la plus proche. (Wiancko, Arny et Salmon 1921).
- la méthode de la marche du cavalier (Tedin 1931). On utilise des micro-parcelles, chaque variété est observée sur plusieurs micro-parcelles qui se déduisent l'une de l'autre par le déplacement du cavalier aux échecs.

Plan en marche de cavalier avec comparaison au témoin



La méthode de Fisher

Les trois principes de Fisher qui aboutiront à la publication de son livre "The design of experiments" (1935).

- Randomisation : introduire du hasard dans l'expérience pour garantir l'équitabilité ; chaque traitement a autant de chances d'être favorisé ou défavorisé .
  Par ailleurs la randomisation valide l'analyse statistique.
- Réplication : Chaque traitement doit être examiné plusieurs fois pour apprécier la variabilité. Pas d'utilité en numérique : pas de variabilité.
- Contrôle local : le plan doit répartir au mieux les différentes occurrences d'un même traitement.

La méthode de Fisher

Les trois principes de Fisher qui aboutiront à la publication de son livre "The design of experiments" (1935).

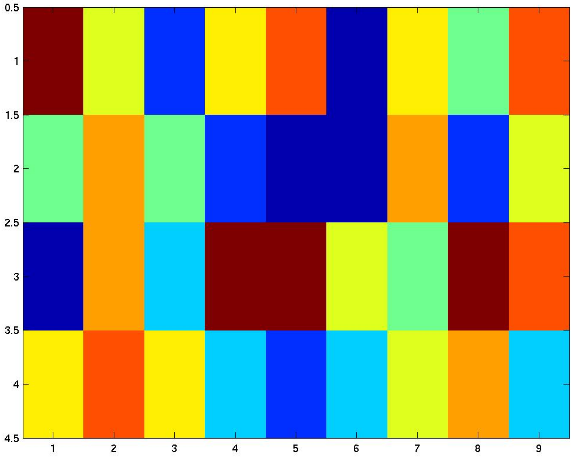
- Randomisation : introduire du hasard dans l'expérience pour garantir l'équitabilité ; chaque traitement a autant de chances d'être favorisé ou défavorisé .
  Par ailleurs la randomisation valide l'analyse statistique.
- Réplication : Chaque traitement doit être examiné plusieurs fois pour apprécier la variabilité. Pas d'utilité en numérique : pas de variabilité.
- Contrôle local : le plan doit répartir au mieux les différentes occurrences d'un même traitement.

La méthode de Fisher

Les trois principes de Fisher qui aboutiront à la publication de son livre "The design of experiments" (1935).

- Randomisation : introduire du hasard dans l'expérience pour garantir l'équitabilité ; chaque traitement a autant de chances d'être favorisé ou défavorisé .
  Par ailleurs la randomisation valide l'analyse statistique.
- Réplication : Chaque traitement doit être examiné plusieurs fois pour apprécier la variabilité. Pas d'utilité en numérique : pas de variabilité.
- Contrôle local : le plan doit répartir au mieux les différentes occurrences d'un même traitement.

Plan en randomisation totale



Exemple :

On veut comparer 4 peintures sur 4 maisons carrées ayant 4 façades de mêmes expositions N, S, E, O. Le facteur d'intérêt est le facteur peinture, les facteurs maison et orientation étant des facteurs à contrôler type bloc. Si on veut équilibrer les relations du premier facteur (peinture) avec chacun des deux autres (maison et orientation), on peut être amené (avant randomisation) à utiliser la répartition suivante :

	Orientation			
Maison	Nord	Sud	Est	Ouest
1	peinture 1	peinture 2	peinture 3	peinture 4
2	peinture 2	peinture 3	peinture 4	peinture 1
3	peinture 3	peinture 4	peinture 1	peinture 2
4	peinture 4	peinture 1	peinture 2	peinture 3

On appelle carré latin, un plan comprenant n^2 unités, réparties en n lignes et n colonnes sur lequel on place un traitement à n valeurs (n lettres latines) de façon que chaque lettre apparaisse une fois et une seule en ligne et en colonne.

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

Le sudoku est un exemple de carré latin avec restrictions

Carrés gréco-latins

La propriété de carré latin peut être enrichie en définissant les carrés greco-latins. Dans le même tableau, on ajoute dans chaque case une lettre parmi n lettres grecques. Chaque lettre grecque doit apparaître une fois et une seule avec une lettre latine. Voici un exemple $4x4$.

De ces plans ont été déduits les **hypercubes latins** dont nous allons parler plus tard.

Plans fractionnaires

La méthode des plans d'expériences a été adaptée à des expériences industrielles dans les années 1940-1950 par Box et ses co-auteurs.
Une des parties qui va nous intéresser le plus est les **plans fractionnaires**. On considère, par exemple, l'émission de polluants d'un moteur diesel dont le réglage dépend d'un très grand nombre de paramètres que l'on appellera **facteurs** par exemple : le couple, le régime, les instants d'injection, les masses injectées, le taux de recyclage etc...

Chaque mesure est très coûteuse. Pas possible de faire toutes les mesures dans toutes les situations.
On va supposer de plus, pour simplifier, que chaque facteur ne prend que deux valeurs.

On va appeler "traitement" une combinaison des p paramètres. Un calcul élémentaire montre qu'il y a 2^p traitements .

Plan factoriel complet

On appelle **plan complet** le plan obtenu en réalisant une fois et une seule chacun des traitements .

En général ce plan est trop grand par rapport aux moyens financiers et on va chercher un sous ensemble algébrique (il est construit avec des techniques d'algèbre) du plan complet qui ait de bonnes propriétés.

Un tel plan aura pour taille la moitié, le quart, le huitième etc... du plan complet. D'où le nom de **plan fractionnaire** .

Le plus difficile est de définir ce que veut dire "qui ait de bonnes propriétés". Pour cela il faut introduire une notion relativement technique : **les interactions multiples**.

Interactions multiples

E : espaces des réponses aux différents traitements.
Exemple : $p = 2$ facteurs, 4 traitements qui correspondent aux 4 combinaisons possibles.
Ensemble des réponses : 4 réponses rangées dans un tableau $2x2$.

$$E \text{ est l'ensemble des } \begin{array}{|c|c|} \hline x & y \\ \hline z & t \\ \hline \end{array}$$

Dans cet espace nous allons définir **inductivement**

- L'espace des fonctions constantes :

$$C : \begin{array}{|c|c|} \hline x & x \\ \hline x & x \\ \hline \end{array}.$$

• L'espace des fonctions qui ne dépendent que du premier facteur mais qui sont orthogonales à l'espace précédent. On part de l'espace des réponses de la forme :

$$F1 : \begin{array}{|c|c|}\hline y & y \\ \hline z & z \\ \hline \end{array} \longrightarrow \text{ orthogonalisation } \longrightarrow \begin{array}{|c|c|}\hline y & y \\ \hline -y & -y \\ \hline \end{array}$$

• L'espace des fonctions qui ne dépendent que du second facteur mais qui sont orthogonales à l'espace des réponses constantes.

$$F2 : \begin{array}{|c|c|} \hline y & -y \\ \hline y & -y \\ \hline \end{array}$$

• L'interaction entre les deux facteurs est définie comme l'ensemble des réponses aux deux facteurs qui sont orthogonales aux trois espaces précédents

$$INT : \begin{array}{|c|c|} \hline y & -y \\ \hline -y & y \\ \hline \end{array}$$

PUB

PUB PUB PUB PUB

Au prix de quelques formules, cette définition se généralise à un nombre quelconque de facteurs, voir

PUB PUB PUB PUB


SCIENCES SUP
Jean-Marc Azaïs
Jean-Marc Bardet
LE MODÈLE LINÉAIRE PAR L'EXEMPLE
Régression, analyse de la variance et plans d'expériences illustrés avec R, SAS et Splus
Y a-t-il une différence de goût entre diverses boissons au cola ? Comment faire un béton sans bulle ? Comment évolue le prix d'une voiture d'occasion en fonction de son modèle et de son âge ?
Pour répondre à ces questions, cet ouvrage présente des variations sur le thème du modèle linéaire (régression linéaire, analyse de la variance, modèles mixtes et plans d'expériences). Chaque notion est introduite et illustrée à l'aide de nombreux exemples utilisant des données réelles. Les programmes permettant leurs traitements avec les logiciels R, SAS et Splus sont présentés, tout comme les résultats numériques et graphiques. Des exercices à la fin de chaque chapitre permettent à l'étudiant de s'entraîner. Les corrigés des exercices ainsi que les bases de données utilisées sont disponibles sur le site web www.dunod.com.
Ne nécessitant pas de connaissances approfondies dans le domaine, ce manuel est destiné aussi bien aux étudiants en Masters de mathématiques appliquées, de biostatistique, d'économétrie, de chimie, qu'aux élèves en école de commerce ou d'agronomie.
JEAN-MARC AZAÏS est professeur de mathématiques appliquées à Toulouse 3.
JEAN-MARC BARDET est professeur de mathématiques appliquées à Paris 1
MATHÉMATIQUES
PHYSIQUE
CHIMIE
SCIENCES DE L'INGÉNIEUR
INFORMATIQUE
SCIENCES DE LA VIE
SCIENCES DE LA TERRE
1er cycle 2e cycle 3e cycle
1 2 3 4 5 6 7 8
LICENCE MASTER DOCTORAT
ISBN 2 10 049559 3
www.dunod.com
DUNOD
J.-M. AZAÏS
J.-M. BARDET
COURS
LE MODÈLE LINÉAIRE PAR L'EXEMPLE
SCIENCES SUP
Cours et exercices corrigés
Master • Écoles d'ingénieurs
LE MODÈLE LINÉAIRE PAR L'EXEMPLE
Régression, analyse de la variance et plans d'expériences illustrés avec R, SAS et Splus
Compléments sur le web
Jean-Marc Azaïs
Jean-Marc Bardet
DUNOD

L'interaction

On aboutit au **résumé** suivant

- l'interaction se définit comme la conjonction particulièrement heureuse ou malheureuse de certains niveaux de certains facteurs. On la note A*B, A*B*C....
- On parle d'interaction double, triple, etc...
- Dans le cas de facteurs à deux niveaux codés -1 +1, l'interaction est engendrée par le produit des facteurs. A B engendre A*B, ABC engendre A*B*C etc...
- Tous ces générateurs sont **orthogonaux**

L'interaction

On aboutit au **résumé** suivant

- l'interaction se définit comme la conjonction particulièrement heureuse ou malheureuse de certains niveaux de certains facteurs. On la note A*B, A*B*C....
- On parle d'interaction double, triple, etc...
- Dans le cas de facteurs à deux niveaux codés -1 +1, l'interaction est engendrée par le produit des facteurs. A B engendre A*B, ABC engendre A*B*C etc...
- Tous ces générateurs sont **orthogonaux**

L'interaction

On aboutit au **résumé** suivant

- l'interaction se définit comme la conjonction particulièrement heureuse ou malheureuse de certains niveaux de certains facteurs. On la note A*B, A*B*C....
- On parle d'interaction double, triple, etc...
- Dans le cas de facteurs à deux niveaux codés -1 +1, l'interaction est engendrée par le produit des facteurs. A B engendre A*B, ABC engendre A*B*C etc...
- Tous ces générateurs sont **orthogonaux**

L'interaction

On aboutit au **résumé** suivant

- l'interaction se définit comme la conjonction particulièrement heureuse ou malheureuse de certains niveaux de certains facteurs. On la note A*B, A*B*C....
- On parle d'interaction double, triple, etc...
- Dans le cas de facteurs à deux niveaux codés -1 +1, l'interaction est engendrée par le produit des facteurs. A B engendre A*B, ABC engendre A*B*C etc...
- Tous ces générateurs sont **orthogonaux**

Le principe fondamental des plans fractionnaires

L'idée clef de la méthode est que

- les effet principaux ont plus de chance d'être plus importants que les interactions doubles
- les interactions doubles ont plus de chance d'etre plus importantes que les interactions triples
- etc...

Le principe fondamental des plans fractionnaires

L'idée clef de la méthode est que

- les effet principaux ont plus de chance d'être plus importants que les interactions doubles
- les interactions doubles ont plus de chance d'etre plus importantes que les interactions triples
- etc...

Le principe fondamental des plans fractionnaires

L'idée clef de la méthode est que

- les effet principaux ont plus de chance d'être plus importants que les interactions doubles
- les interactions doubles ont plus de chance d'etre plus importantes que les interactions triples
- etc...

Premier plan fractionnaire

Trois facteurs : Plan complet avec générateurs :

moy	*A*	*B*	*C*	*AB*	*AC*	*BC*	*ABC*	traitement
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	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$	$(+1, +1, +1)$

Budget : nous décidons de ne conserver que 4 unités. Par exemple, les 4 unités pour lesquelles $A.B.C = 1$, ce qui définit un "demi plan" ou "fraction".

Premier plan fractionnaire

trois facteurs : plan factoriel complet avec différents générateurs :

moy	A	B	C	AB	AC	BC	ABC	traitement
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	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$	$(+1,+1,+1)$

Quelles sont les propriétés de la fraction ?

On est amené à résoudre le système de quatre équations :

$$
\begin{aligned}
f(--+) &= e(1) + e(ABC) - e(A) - e(BC) - e(B) - e(AC) \\
&\quad + e(C) + e(A \cdot B) \\
f(-+-) &= e(1) + e(ABC) - e(A) - e(BC) + e(B) + e(AC) \\
&\quad - e(C) - e(AB) \\
f(+--) &= e(1) + e(ABC) + e(A) + e(BC) - e(B) - e(AC) \\
&\quad - e(C) - e(AB) \\
f(+++) &= e(1) + e(ABC) + e(A) + e(BC) + e(B) + e(AC) \\
&\quad + e(C) + e(AB)
\end{aligned}
$$

Les confusions

Dans un plan fractionnaire donné par une clef
par exemple ABC =1 il y a des confusions entre interactions et effets principaux qui se déduisent de la clef.
Comme A,B,C valent ± 1

$$A^2 = B^2 = C^2 = 1$$

et

$$1 = ABC \Rrightarrow A = BC \text{ par exemple}$$

Méthode des facteurs de base

Plan $2^{p-q}(p > q)$: un plan à 2^{p-q} unités pour p facteurs . Taux de fraction : 2^{-q}.

Construction
(1) Plan factoriel complet pour les $(p-q)$ premiers facteurs
facteurs de base . $\Rightarrow 2^{p-q}$ unités.
(2) Définir ensuite les valeurs des q autres facteurs par q clefs.

Ex Plan 2^{6-2} pour les facteurs A, B, C, D, E, F
On construit le plan complet pour A, B, C, D ce qui donne bien 16 unités. On construit les valeur de E et F par

$$E = ABC \quad ; \quad F = BCD$$

Calcul des confusions

Des deux clefs

$$E = ABC \quad ; \quad F = BCD$$

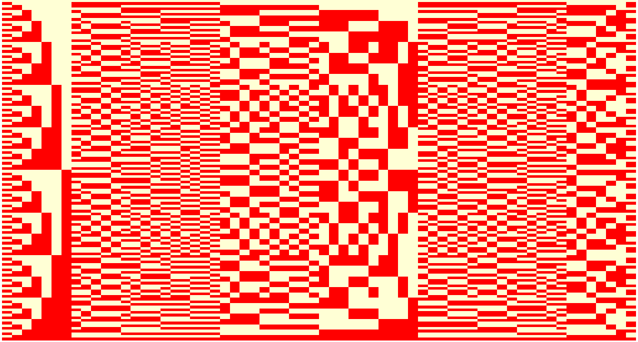
on déduit par combinaison **l'alias** de 1 :

$$1 = ABCE = BCDF = ADEF$$

D'où le calcul de **l'alias** d'un terme
(effet principal ou interaction) :

$$F = ABCDEF = BCD = ADE \quad ; \quad AB = CE = ACDF = BDEF.$$

Plan 2^{64-57}



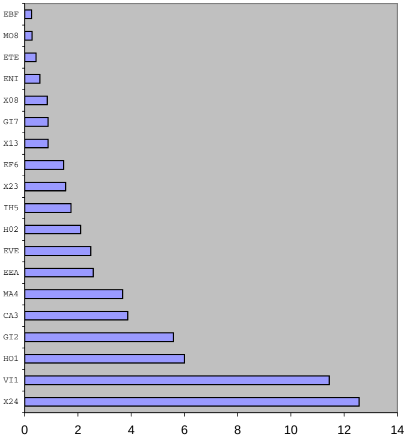
Voici un exemple d'analyse du plan :

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	27	1.528363E-13	5.660604E-15	2.21	0.0025
Error	100	2.565456E-13	2.565456E-15		
Total	127	4.093819E-13			

R2	CV	Root MSE	REP Mean
0.37	82.14	5.06503E-8	6.16636E-8

Source	DF	Mean Square	F Value	Pr > F
H02	1	5.379512E-15	2.10	0.1507
SA9	1	2.745152E-16	0.11	0.7443
MA4	1	9.436798E-15	3.68	0.0580 *
Re1	1	7.441283E-17	0.03	0.8651
IH5	1	4.474904E-15	1.74	0.1896
VI1	1	2.936159E-14	11.44	0.0010 **
MO8	1	7.302005E-16	0.28	0.5949
X08	1	2.175403E-15	0.85	0.3593
DE2	1	6.069954E-19	0.00	0.9878
HO1	1	1.538497E-14	6.00	0.0161 **
DE3	1	1.356442E-16	0.05	0.8186
EVE	1	6.369129E-15	2.48	0.1183

etc..


EBF
MO8
ETE
ENI
X08
GI7
X13
EF6
X23
IH5
H02
EVE
EEA
MA4
CA3
GI2
HO1
VI1
X24
0
2
4
6
8
10
12
14

La Résolution est la plus petite longueur d'un mot dans l'alias de 1. Si, par exemple, on a :

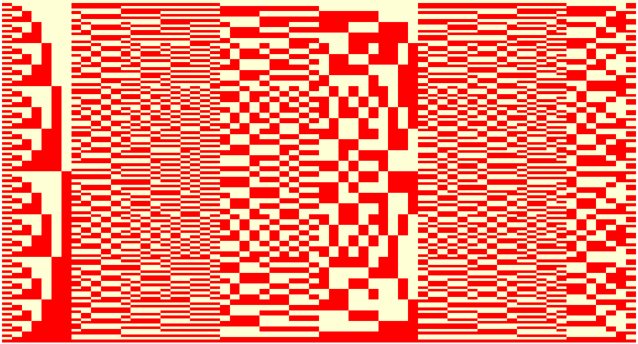
$$1 = ABCE = \mathbf{BCF} = \mathbf{AEF},$$

la résolution est **III**. Dans la théorie statistique des plans d'expériences on suppose souvent qu'il n'y a pas d'interaction triple et plus. Dans ce cas :

- Une résolution V correspond à un plan "haut de game"
- Une résolution IV correspond à un plan "milieu de game"
- Une résolution III est le minimum acceptable.

- Un plan de résolution ρ a des marginales d'ordre $\rho - 1$ équilibrées.
  Par exemple, soit un plan de résolution *IV* à 128 unités. Soit un triplet quelconque de variables, par exemple, B, D, E qui peut prendre 8 valeurs différentes. Alors chacune est vue $128/8 = 16$ fois.
- Dans un plan de résolution V, il n'y a aucune confusion entre effets principaux et interaction doubles.

Plan 2^{64-57}



Dans des plans sur certains codes numériques on trouve parfois des interactions élevées "significatives".

Source	DF	F Value
model	3	13315.5
aero	2	138684
visi	3	395434
hum	2	48812.7
dist	3	932580
model*aero	6	313.70
⋮	⋮	⋮
aero*visi*hum*dist	36	207.18

Pour de telles situations, des résolutions de VI et plus sont nécessaires et il n'y a pas de méthodes de construction standard au delà d'un vingtaine de facteurs : **Problème ouvert**

Pour des nombre de facteurs < 20 il existe des programmes

- proc Factex de SAS
- commande fractfact de MATLAB X = fracfact('a b c d abc bcd acd')
- et il existe des tables d'emploi aisé
- il existe des algorithmes généraux de construction de plans de résolution III et IV.

Nombre unités	Plans fractionnaire pour p facteurs à deux niveaux — Nombre de facteurs p: 3	4	5	6	7	8	9	10	11
4	III 2^{3-1} C=A.B	/	/	/	/	/	/	/	/
8	Plan Complet	IV 2^{4-1} D=ABC	III 2^{5-2} D=A.B E=A.C	III 2^{6-3} D=A.B E=A.C F=B.C	III 2^{7-4} D=A.B E=A.C F=B.C G=A.B.C	/	/	/	/
16		Plan Complet	V 2^{5-1} E=A.B.C.D	IV 2^{6-2} E=A.B.C F=B.C.D	IV 2^{7-3} E=A.B.C F = B.C.D G = A.C.D	IV 2^{8-4} E=B.C.D F=A.C.D G=A.B.C H=A.B.D	III 2^{9-5} E=A.B.C F=B.C.D G=A.C.D H=A.B.D I=A.B.C.D	III 2^{10-6} E=A.B.C F=B.C.D G=A.C.D H=A.B.D I=A.B.C.D J=A.B	III 2^{11-7} E=A.B.C F=B.C.D G=A.C.D H=A.B.D I=A.B.C.D J=A.B K =A.C
32			Plan Complet	VI 2^{6-1} F=A.B C.D.E	IV 2^{7-2} F=A.B.C.D G=A.B.D.E	IV 2^{8-3} F=A.B.C G=A.B.D H=B.C.D.E	IV 2^{9-4} F=B.C.D.E G=A.C.D.E H=A.B.D.E I=A.B.C.E	IV 2^{10-5} F=A.B.C.D G=A.C.D.E H=A.B.D.E I=A.C.D.E J=B.C.D.E	IV 2^{11-6} F=A.B.C G=B.C.D H=C.D.E I=A.C.D J=A.D.E K =B.D.E
64				Plan Complet	VII 2^{7-1} G=A.B.C D.E.F	V 2^{8-2} G=A.B.C.D H=A.B.E.F	IV 2^{9-3} G=A.B.C.D H=A.C.E.F I=C.D.E.F	IV 2^{10-4} G=B.C.D.F H=A.C.D.E I=A.B.D.E J=A.B.C.E	IV 2^{11-5} G=C.D.E H=A.B.C.D I=A.B.F J=B.D.E.F K=A.D.E.F
128					Plan Complet	VIII 2^{8-1} H=A.B.C. D.E.F.G	VI 2^{9-2} H=A.C.D F.G I=B.C.E.F.G	V 2^{10-3} H=A.B.C.G I=B.C.D.E J=A.C.D.F	V 2^{11-4} H=A.B.C.G I=B.C.D.E J=A.C.D.F K=A.B.C. D.E.F.G

Tab. 1 – Tableau de plans fractionnaires de minimum d'aberration, avec, en ligne, le nombre d'unités 2^{p-q} et en colonne le nombre de facteurs. Une solution est donnée lorsqu'il existe un plan

La méthode des plans fractionnaires telle qu'on l'a décrite fonctionne pour

- des facteurs à 4,8,etc.. niveaux. Il suffit de coder en binaire.
- Des facteurs qui on tous un nombre de niveaux qui est une puissance d'un même nombre premier, par exemple 3,9. Il faut alors coder les niveaux par des racines $p$ièmes de l'unité dans le corps $\mathbb{C}$ des nombres complexes.
- On ne peut pas mélanger simplement des facteurs à 3 niveaux et à 2 niveaux, par exemple

Dans le cas le plus complexe, il faut résoudre des équations dans des corps finis et c'est l'affaire de spécialistes mais cela existe.

Remarque : certains liens existent :

- Un carré latin de taille n peut être vu comme un plan fractionnaire n^{3-1}.
- Un carré Greco-latin de taille n est un plan n^{4-2}

La prochaine conférence traitera de facteurs quantitatifs qui varient librement dans un intervalle. Nous soulignerons a la fin les liaisons entre les deux situations.