

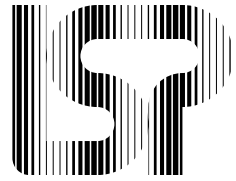
UNIVERSITE  
PAUL  
SABATIER



TOULOUSE III

PUBLICATIONS DU LABORATOIRE  
DE  
STATISTIQUE ET PROBABILITÉS

---



# Composantes de la variabilité Plans d'expériences

Jean-Marc Azaïs

---

Laboratoire de Statistique et Probabilités — UMR CNRS C55830  
Université Paul Sabatier — 31062 – Toulouse cedex 4.



# Chapitre 1

## Modèles mixtes

### 1 Analyse de la variance multivariante

Dans cette section nous présentons l'analyse de la variance multivariante dans sa présentation classique. En fait ses modèles peuvent se traiter comme un cas particulier des modèles mixtes que nous allons décrire dans la suite de cet ouvrage.

L'analyse de la variance, la régression et le modèle linéaire en général, ont pour but d'expliquer **une** variable quantitative. Cependant, dans de nombreux cas, l'observation est multiple : on peut par exemple mesurer la longueur et la largeur d'une feuille, le rendement d'une réaction chimique en plusieurs instants, etc... Une analyse variable par variable est toujours possible, elle donne des estimations qui sont parfaitement exploitables. Pour ce qui est des tests, la situation est plus complexe : chaque analyse uni-variable donne un test différent pour une hypothèse comme par exemple la nullité de l'effet d'un facteur. Ces tests sont peu puissants car ils n'utilisent que l'information d'une seule variable et ils peuvent même être contradictoires. Il est donc très souhaitable de pouvoir regrouper les tests sur les différentes variables.

*Exemple :* Pour apprécier et comparer le degré de pollution de rivières, on dose différents métaux lourds dans les viscères de truites. On dispose par exemple de 6 observations : Plomb dans le foie, plomb dans l'intestin, plomb dans le colon, Mercure dans le foie, mercure dans l'intestin, mercure dans le colon.

On suppose que l'on observe plusieurs truites par rivières et on note  $Y_{ij}$  l'observation sur la truite  $j$  de la rivière  $i$ . On se trouve dans une situation d'analyse de la variance à un facteur **à ceci près que l'observation se trouve dans  $\mathbb{R}^6$** . On pose

$$Y_{ij} = t_i + \epsilon_{ij} \ ; \ t_i \in \mathbb{R}^6 \ ; \ \epsilon_{ij} \sim N(0, \Sigma) \ ; \ i = 1, I \ ; \ j = 1, n_i,$$

où  $\Sigma$  est une matrice 6x6 définie positive.

Comme nous l'avons déjà souligné, l'estimation peut parfaitement se faire variable par variable, par contre le test de la nullité "l'effet rivière" par exemple aura 6 réponses en fonction de la variable utilisée.

Pour obtenir un test unique, nous allons paraphraser les formules de l'analyse de la variance uni-variable en essayant de les généraliser au cas vectoriel.

-On peut toujours estimer  $t_i$  par  $Y_i$  qui est le vecteur des moyennes. On définit ainsi

le vecteur des résidus associé à l'unité  $i, j$ .

$$\hat{\epsilon}_{ij} = Y_{ij} - Y_{i.}$$

Comme il s'agit toujours d'un vecteur de taille 6 ( que l'on notera maintenant  $L$  pour généraliser), la somme des carrés est remplacé par une matrice  $L \times L$ . Si on note en exposant les coordonnées d'un vecteur ou d'une matrice, on a

$$SCR^{k,l} = \sum_{i,j} \hat{\epsilon}_{ij}^k \hat{\epsilon}_{ij}^l, \quad k = 1, L ; l = 1, L.$$

Dans le cas uni variable cette quantité suivait une loi  $\sigma^2 \chi^2(n-I)$  ( $n = I \times J$  est le nombre d'unités). Dans le même état d'esprit on définit

**Définition 1 (loi de Whishart de paramètres  $d$  et  $\Sigma$ )** Soient  $d$  vecteurs indépendants  $X_1, \dots, X_d$  de loi  $N(0, \Sigma)$ . On définit la matrice  $M$  des sommes de carrés et produits par

$$M = \sum_{i=1,d} X_i(X_i)'$$

où le  $'$  désigne la transposée. Cette matrice est composée de sommes de carrés qui suivent de lois  $\sigma^2 \chi^2(d)$  sur la diagonale et de sommes de produits à l'extérieur de la diagonale. Par définition elle suit la loi de Whishart de paramètres  $d$  et  $\Sigma$ .

Par les mêmes techniques qu'en uni-variable, on montre que

$$SCR \text{ suit un loi de Whishart } N - I, \Sigma.$$

On définit également la somme de carrés et produits associée au facteur :

$$SCF := \sum_{i,j} (Y_{i.} - Y_{..})(Y_{i.} - Y_{..})'$$

Sous l'hypothèse  $H_0$  d'absence d'effet du facteur (rivière dans notre exemple), on montre que

$$SCF \text{ suit une loi de Whishart } I - 1, \Sigma.$$

Il reste à construire le test. Rappelons qu'en uni-variable, ce test est basé sur le rapport

$$\hat{F} := \frac{SCF/(I-1)}{SCR/(N-I)}$$

qui suit la loi de Fisher  $F_{(I-1),(N-I)}$ . En multivariable, d'un certain coté, la situation est analogue puisque  $SCR/(N-I)$  est un estimateur de  $\Sigma$  et d'un autre coté, différente dans le sens qu'il n'est pas possible de faire le quotient de deux matrices.

On pose  $W := SCR/(N-I)$  (comme within) matrice des variations intra-facteur et  $B := SCF/(I-1)$  (comme between) matrice de variations inter-facteur. Les tests en multivariable sont multiples et basés sur les statistiques suivantes

- le  $\lambda$  de Wilks :  $\frac{|W|}{|B+W|}$  qui donne le test du rapport de vraisemblance,
- la trace de Pillai :  $Tr(B((B+W)^{-1}))$ ,
- la trace de Hotelling-Lawley :  $Tr(BW^{-1})$ ,
- la plus grande valeur propre de Roy : PGVP ( $W^{-1}B$ ).

Fort heureusement, dans la plupart des cas ces tests donnent des résultats cohérents. Les lois connues et complexes sous  $H_0$  ont de très bonnes approximation par des lois de Fisher qui sont utilisées le plus souvent par les logiciels pour calculer les niveaux de signification.

### Mise en œuvre SAS (cas équilibré)

```
Proc Glm data=sasuser.pollution;
Class riviere;
model Hgf Hgi Hgc Pbf Pbi Pbc = riviere;
means riviere/Tukey;
Manova h=riviere; run;
```

Ce modèle d'analyse de la variance multivariable peut apparaître comme un cas particulier de modèle mixte que nous allons présenter dans les section suivantes.

## 2 Modèles à effets aléatoires et mixtes

### 2.1 Facteurs croisés et hiérarchisés

Quand on considère deux facteurs en situation d'analyse de la variance, deux situations peuvent se rencontrer :

- Chacun d'eux a un sens indépendamment de l'autre, c'est le cas le plus courant, les facteurs sont dits **croisés**. Exemple de facteurs en général croisés :

variété	×	lieu
machine	×	opérateur
concentration	×	température (réaction chimique)

- Un des facteur, par exemple  $B$  ne signifie rien de concret tant que l' on ne connaît pas l'indice associé de l'autre facteur ( $A$ ). Dans ce cas le facteur  $B$  est **hiérarchisé** au facteur  $A$ . C'est le plus souvent le cas quand le facteur  $B$  correspond à une numérotation à l'intérieur des niveaux du facteur  $A$ . Considérons par exemple plusieurs portées de lapereau, et supposons que dans chaque portée, on numérote chaque lapereau de 1 au nombre de lapereaux dans la portée. Il ne peut y avoir d'effet propre du numéro de lapereau car il n'y a aucun rapport entre les lapereaux qui porte un même numéro.

Exemple de facteurs le plus souvent hiérarchisés

famille génétique / $n^o$ de descendant
lapine / $n^o$ de portée / $n^o$ du lapin dans la portée.

**Remarque** Une autre alternative consisterait a ne pas numéroter le facteur hiérarchisé a l'intérieur des niveaux du facteur hiérarchisant : dans l'exemple des lapereau, si le première lapine donne naissance par exemple à 4 lapins, le premier lapereau de la seconde portée serait numéroté 5. C'est alors un troisième type de relation entre deux facteurs : l'information de l'un (lapine) est alors comprise dans l'information de l'autre. En fait, pour des raisons d'organisation des logiciels statistiques cette second approche ne donne pas des décompositions satisfaisantes : le faire sur un exemple pour s'en convaincre. En conséquence on préfère souvent la première approche.

Quand on est en présence de facteurs hiérarchisés, il faut prendre soin de le déclarer avec la syntaxe propre au logiciel que l'on utilise. En effet, il ne faut pas, dans ce cas, définir un effet propre du facteur hiérarchisé. La décomposition est différente de celle du modèle croisé dans le sens ou ce qui serait l'effet principal du facteur hiérarchisé est incorporé à l'interaction.

Plus précisément on écrit

$$Y_{ijk} = \mu + \alpha_i + \gamma_{ij} + \epsilon_{ijk}$$

Où  $\gamma_{ij}$  est l'effet du facteur hiérarchisé, avec les contraintes

- i.  $\sum_{i=1,I} \alpha_i = 0$
- ii. pour tout  $i$   $\sum_{j=1,J} \gamma_{ij} = 0$ .

## 2.2 Modèles mixtes équirépétés

Exemple : On désire comparer les aptitudes de 4 firmes à produire des insecticides efficaces. Chaque firme produit de nombreux insecticides mais on échantillonne exactement deux produits numérotés (1 et 2) par firme. Pour étudier les insecticides on utilise 24 boîtes contenant 400 moustiques chacune. Chaque produit est introduit dans 3 boîtes prises au hasard parmi les 24 et on compte le nombre de moustiques survivants au bout d'un certain temps. On considère que le nombre 400 est suffisamment grand pour que l'aspect binomial discret du problème puisse être oublié. Cela permet d'éviter d'utiliser un modèle linéaire généralisé à effets aléatoires. On reconnaît une structure où l'effet insecticide est hiérarchisé au facteur firme. Comme les produits ont été échantillonnés le facteur produit est aléatoire.

On veut répondre aux questions : Y a-t'il une différence entre les firmes ? Quelle est la variabilité la plus importante : celle relative au choix du produit ou la variabilité résiduelle ?

Pour l'instant nous allons voir comment répondre à la seconde question. On note  $Y_{ij,k}$  l'observation dans la  $i$ ème boîte du  $j$ ème produit de la firme  $k$ ,  $i = 1, I(4)$  ;  $j = 1, J(2)$  ;  $k = 1, K(3)$  et on pose le modèle

$$Y_{ijk} = \mu + f_i + P_{ij} + \epsilon_{ijk}$$

où les majuscules désignent des effets aléatoires. On suppose qu'ils sont tous indépendants, à l'intérieur d'un effet et d'un effet à l'autre et que les  $P_{ij}$ ,  $\epsilon_{ijk}$  sont centrés gaussiens de variances respectivement  $\sigma_P^2$  et  $\sigma^2$ . L'importance relative des effets va s'apprécier par l'estimation des différentes variances. Ici on est dans un cas équirépété, l'estimation en sera très simple. Il existe un estimateur qui est uniformément meilleur par des techniques d'exhaustivité (pour une condition exacte pour qu'il en soit ainsi voir Coursol (1980)). On définit les sommes de carrés résiduelle  $SCR$  et associée à l'effet produit  $SCP$ . Par utilisation de l'orthogonalité du modèle il est facile de vérifier que

$$SCR = \sum_{ijk} (Y_{ijk} - Y_{ij.})^2$$

$$SCP = \sum_{ijk} (Y_{ij.} - Y_{i..})^2.$$

On laisse à titre d'exercice de montrer que

$$E(SCP) = I(J-1)K\sigma_P^2 + I(J-1)\sigma^2 \quad (1.1)$$

$$E(SCR) = IJ(K-1)\sigma^2 \quad (1.2)$$

On obtient un système triangulaire donc inversible qui donne un estimateur de  $\sigma_P^2$ ,  $\sigma^2$  en identifiant les sommes de carrés à leurs espérances, c'est à dire en résolvant le système

$$SCP = I(J-1)K\hat{\sigma}_P^2 + I(J-1)\hat{\sigma}^2 \quad (1.3)$$

$$SCR = IJ(K-1)\hat{\sigma}^2 \quad (1.4)$$

Cela correspond à une méthode de moments. Comme le cas équiréparté reste un cas particulier, nous ne détaillerons pas, mais on montre (Coursol 1980) que dans ce cas les estimateurs ci dessus sont optimaux parmi les estimateurs sans biais.

### 2.3 Modèles mixtes généraux

Première remarque : la méthode précédente reste toujours applicable. Si on utilise les sommes de carrés de type III pour l'identification, elle porte le nom de la **Méthode de Henderson III**. Le calcul des coefficients dans (1.1)(1.2) est plus complexe, mais des expressions matricielles peuvent être établies par les techniques que nous allons développer plus loin, expressions qui peuvent être calculées par l'ordinateur. Cette méthode est parfois très acceptable surtout pour les dispositifs peu déséquilibrés, et dans tous les cas elle fournit de bon point de départ pour les méthodes itératives que nous allons développer plus loin.

Au départ un modèle mixte est la donnée d'un vecteur  $Y$  qui peut s'écrire

$$Y = X\theta + Z_1\beta_1 + \dots + Z_{k-1}\beta_{k-1} + \epsilon$$

où  $\theta$  est un vecteur de paramètres inconnus, les matrices  $X, Z_1, \dots, Z_k$  sont des matrices connues de dimension convenables et  $\beta_1, \dots, \beta_k$  sont de vecteurs indépendants gaussiens formés de coordonnées indépendantes et de même variance (on note  $\gamma_i$  la variance commune des coordonnées de  $\beta_i$ ). Le modèle est mixte dans le sens où il regroupe de effets déterministes  $\theta$  et des effets aléatoires. De la formule ci dessus on déduit facilement que

$$E(Y) = X\theta$$

$$\text{Var}(Y) = V_\gamma := \sum_{i=1,k} \gamma_i Z_i Z_i'$$

en posant  $V_k = Id$ . C'est cette forme que l'on va poser comme définition.

**Définition 2 (modèles mixtes)** *Un vecteur aléatoire gaussien  $Y$  de taille  $n$  est dit suivre un modèle mixte statistique si*

$$E(Y) = X\theta$$

$$\text{Var}(Y) = V_\gamma := \sum_{i=1,k} \gamma_i V_i.$$

où  $X, V_1, \dots, V_k$  sont de matrices connues,  $\theta$  est un vecteur de paramètres inconnus variant dans  $\mathbb{R}^p$ ,  $\gamma := (\gamma_1, \dots, \gamma_k)$  varie dans l'ensemble  $S := \{\gamma : V_\gamma > 0\}$ .

Ce modèle est noté :  $Y \sim MM(X, V_1, \dots, V_k)$ . Les  $(\gamma_1, \dots, \gamma_k)$  sont appelées composantes de la variance. On pose  $Y = X\theta + \epsilon$  avec  $\text{Var}(\epsilon) = V_\gamma$ .

## 2.4 Estimation des effets fixes

Pour simplifier on supposera  $X$  de plein rang, on a vu que l'on pouvait s'y ramener le plus souvent. Si  $\gamma$  est connu, le modèle mixte se ramène à un modèle linéaire ordinaire et l'estimateur optimal parmi les estimateurs linéaires sans biais est

$$\hat{\theta} = (X'V_\gamma^{-1}X)^{-1}X'V_\gamma^{-1}Y. \quad (1.5)$$

*Démonstration* : Par diagonalisation il existe une matrice  $T$  symétrique que l'on notera par la suite  $V_\gamma^{-1/2}$  telle que

$$TT = V_\gamma^{-1} \quad ; \quad TV_\gamma T = Id$$

On pose donc  $\tilde{Y} = TY = TX\theta + T\epsilon = \tilde{X}\theta + \tilde{\epsilon}$ . Il est facile de voir que  $Var(\tilde{\epsilon}) = Id$  et que donc  $\tilde{Y}$  suit un modèle linéaire à variance connue. Dans ce modèle l'estimateur optimal est

$$\hat{\theta} = ((\tilde{X})'\tilde{X})^{-1}(\tilde{X})'\tilde{Y} = (X'V_\gamma^{-1}X)^{-1}X'V_\gamma^{-1}Y.$$

□

Notons que l'équation (1.5) est numériquement facile à résoudre même pour des modèles comprenant plusieurs centaines de paramètres.

De la même manière si on considère une hypothèse linéaire sur le vecteur  $\theta$  le test de Fisher dans le modèle  $\tilde{Y} = \tilde{X}\theta + \tilde{\epsilon}$  sera optimal. Quand  $\gamma$  ne sera plus connu mais estimé, on estimera encore les effets fixes en utilisant les équations ci-dessus avec  $\hat{\gamma}$  comme s'il s'agissait de la vraie valeur. Nous verrons plus loin que cela correspond à la solution du maximum de vraisemblance gaussien. En conclusion le problème statistique essentiel dans un modèle mixte est l'estimation des composantes de la variance.

## 2.5 Estimation par MIVQUE dans un modèle mixte

Supposons que nous voulons estimer une composante  $\gamma_i$  de  $\gamma$  ou plus généralement une combinaison linéaire

$$h = \mathbf{h}'\gamma$$

des composantes.  $\mathbf{h} \in \mathbb{R}^k$  donné. On cherche un estimateur

- (i) invariant :  $\hat{h}(Y) = \hat{h}(Y + X\theta)$ ,  $\forall \theta \in \mathbb{R}^p$
- (ii) quadratique :  $\hat{h}(Y) = Y'BY$  où  $B$  est une matrice (il est basé sur des sommes de carrés).
- (iii) sans biais :  $E(\hat{h}(Y)) = h$
- (iv) de faible variance.

**Proposition 1** (1) On peut toujours supposer  $B$  symétrique et on le fera tout le temps.

(2) L'estimateur est invariant ssi  $BX = 0$  ( $B$  est maintenant symétrique)

(3) Si  $Y$  est gaussien et si  $BX = 0$ ,  $Var(Y'BY) = Var(\epsilon' B \epsilon) = 2Tr(BV_\gamma B V_\gamma)$ .

La difficulté principale vient du fait que la variance de  $\hat{h}$  dépend en général de la valeur de  $\gamma$  que l'on cherche à estimer. Il est donc sans espoir (sauf dans les cas équirépétés) de chercher un estimateur optimal. On va donc se donner un **a priori**  $\gamma_0$  et minimiser la variance gaussienne au voisinage de  $\gamma_0$  :  $Tr(BV_{\gamma_0} B V_{\gamma_0})$ .



**Théorème 1** Soit  $\gamma_0$  un a priori. Un estimateur vérifiant (i) (ii) (iii) et de  $\gamma_0$ -variance minimale est appelé MIVQUE (Minimum Variance Quadratic Unbiased Estimator). Il est donné par les équations

$$\hat{h}(Y) = Y'Q'_{\gamma_0}V_{\gamma_0}^{-1} \sum_{i=1}^k \delta_i V_i V_{\gamma_0}^{-1} Q_{\gamma_0} Y$$

avec

- $Q_{\gamma_0} := Id - X(X'V_{\gamma_0}^{-1}X)^{-1}X'V_{\gamma_0}^{-1}$ ,
- $\delta$  solution de  $\delta' M_{\gamma_0} = \mathbf{h}$
- $M_{\gamma_0} = \{Tr(V_i V_{\gamma_0}^{-1} Q_{\gamma_0} V_j V_{\gamma_0}^{-1} Q_{\gamma_0}), i = 1, k, j = 1, k\}$ .

En particulier si la matrice  $M_{\gamma_0}$  est inversible, il existe un MIVQUE de toute composante.

### MIVQUE vectoriel

Supposons  $M_{\gamma_0}$  inversible, comme toute combinaison est estimable, on obtient un MIVQUE de chaque coordonnée  $\gamma_i$ . Pour ce faire, on cherche  $\delta^i$  tel que

$$M_{\gamma_0} \delta^i = f_i$$

où  $f_i$  est le  $i$ ème vecteur de la base canonique (avec un 1 en  $i$ ème position). Cela implique que

$$\{\delta_l^i, i = 1, k, j = 1, k\} = (M_{\gamma_0})^{-1}$$

On définit le vecteur des sommes de carrés  $S = \{Y'Q'_{\gamma_0}V_{\gamma_0}^{-1}V_l V_{\gamma_0}^{-1}Q_{\gamma_0}Y, l = 1, k\}$ . On a alors

$$\hat{\gamma} = (M_{\gamma_0})^{-1} S.$$

## 2.6 Estimation par maximum de vraisemblance restreinte

On suppose que  $X$  est de plein rang,  $Y$  est un vecteur gaussien. La vraisemblance vaut

$$(2\pi)^{-n/2} |V_{\gamma}|^{-1/2} \exp(-1/2 (Y - X\theta)' V_{\gamma}^{-1} (Y - X\theta))$$

On passe à  $-2 \log$  de cette expression et à une constante près on est conduit à maximiser

$$L(\theta, \gamma) := \log(|V_{\gamma}|) + (Y - X\theta)' V_{\gamma}^{-1} (Y - X\theta)$$

Pour trouver les équations du maximum de vraisemblance, on remarque

$$\frac{\partial V_{\gamma}^{-1}}{\partial \gamma_i} = -V_{\gamma}^{-1} \frac{\partial V_{\gamma}}{\partial \gamma_i} V_{\gamma}^{-1}$$

$$\frac{\partial \log |V_{\gamma}|}{\partial \gamma_i} = Tr\left(\frac{\partial V_{\gamma}}{\partial \gamma_i} V_{\gamma}^{-1}\right).$$

On en déduit que

$$\frac{\partial L(\theta, \gamma)}{\partial \theta} = 0 \Leftrightarrow (X'V_{\gamma}^{-1}X)\theta = X'V_{\gamma}^{-1}Y$$

Cette équation est connue sous le nom d'équation de Gauss-Markov. Elle permet de retrouver l'équation(1.5).

$$\frac{\partial L(\theta, \gamma)}{\partial \gamma_i} = 0 \Leftrightarrow \text{Tr}(V_i V_\gamma^{-1}) = (Y - X\theta)' V_\gamma^{-1} V_i V_\gamma^{-1} (Y - X\theta)$$

Cependant, dans beaucoup de cas les simulations montrent que le maximum de vraisemblance est biaisé. Comme la vraie difficulté, comme on l'a vu réside dans l'estimation des composantes de la variance, on va d'une façon "concentrer" la vraisemblance sur cette estimation.

**Définition 3** *On appelle vraisemblance restreinte, la vraisemblance de  $T'Y$  où  $T$  est une matrice quelconque de rang maximal telle que  $T'X = 0$ . Il est facile de voir que  $T'Y \sim N(0, T'V_\gamma T)$  et donc on pose*

$$L_R(\gamma) := \log(|T'V_\gamma T|) + Y'T(T'V_\gamma T)^{-1}T'Y$$

**Proposition 2**  *$L_R$  ne dépend de  $T$  qu'à travers l'ajout d'une constante qui ne change rien et de plus*

$$\frac{\partial L_R(\gamma)}{\partial \gamma_i} = 0 \Leftrightarrow \text{Tr}(V_i V_\gamma^{-1} Q_\gamma) = Y' V_\gamma^{-1} Q_\gamma V_i V_\gamma^{-1} Q_\gamma Y$$

En conséquence en comparant les équations, on constate qu'un point fixe du MIVQUE ( $\hat{\gamma} = \gamma_0$ ) est une solution des équations du maximum de vraisemblance restreinte. L'itération du MIVQUE est donc une façon (parmi d'autres) de résoudre ces équations. Pour plus de détail on pourra consulter Azais Bardin Dhorne (1993).

## 2.7 Tests

Pour ce qui concerne les tests d'hypothèses sur les effets fixes, on réalise des tests de Fisher en supposant que les estimateurs des composantes de la variance sont en fait les valeurs exactes et en utilisant la méthode de la section 2.4.

Le plus souvent les tests sur les effets aléatoires correspondent à la nullité d'une variance : on peut chercher à tester la nullité d'un effet "famille génétique", ou "sujet" par exemple. La première solution consiste à utiliser un test exact **de Fisher**. En effet la nullité d'un effet aléatoire correspond strictement à l'absence d'effet, c'est à dire également à la nullité d'un effet déclaré en effet fixe. En résumé, pour tester la nullité d'une composante de la variance, on peut déclarer l'effet correspondant en fixe et utiliser le test de Fisher correspondant. Sauf dans le cas équilibré (voir Coursol 1980) ce test n'est plus optimal. Mais il est exact dans le sens où son niveau réel est toujours égal au niveau nominal.

L'autre alternative est d'utiliser les tests classiques asymptotiques associés à la méthode du maximum de vraisemblance : le test du rapport de vraisemblance et le test de Wald. Pour les présenter, nous utilisons les notations traditionnelles où  $\theta$  est le paramètre du modèle statistique. Le  $\theta$  de cette partie est donc en fait égal au  $\theta, \gamma$  du modèles mixte dans le cas de la vraisemblance classique, il est égal au seul  $\gamma$  dans le cas de la vraisemblance restreinte.

Dans une expérience qui comprend un grand nombre de répétitions indépendantes et sous des hypothèses de régularité (modèles de vraisemblance réguliers) qui sont vérifiées dans notre cas, on sait que l'estimateur du maximum de vraisemblance est asymptotiquement sans biais, normal et de variance donnée par l'inverse de l'information de Fisher (Dacunha-Castelle & Dufflo, 1983). L'information de Fisher  $I(\theta)$  est donnée par

$$(I(\theta))_{ij} = -E\left(\frac{\partial^2 \log V(\theta)}{\partial \theta_i \partial \theta_j}\right)$$

où  $V$  est la vraisemblance,  $\theta$  le paramètre du modèle. Comme,  $I(\theta)$  tend vers l'infini avec le nombre de répétition, la phrase " asymptotiquement sans biais, normal et de variance donnée par l'inverse de l'information de Fisher" doit se comprendre comme

$$(I(\theta))^{1/2}(\hat{\theta} - \theta) \rightarrow N(0, Id)$$

quand le nombre de répétitions tend vers l'infini.

Une telle propriété reste vraie pour la vraisemblance restreinte et dans des cadres plus généraux, mais sous des conditions très techniques (Jiang 1996). Ces conditions sont en particulier que les composantes de la variance correspondent à des effets aléatoires  $\beta_1 \dots \beta_k$  au sens de notre première définition d'un modèle mixte et que ces composantes soient toutes positives. Dans ces conditions il est possible de construire un test de Wald de l'hypothèse

$$\theta_i = \theta_{i,0}$$

dont la région de rejet est

$$\left| (I(\theta)_{ii}^{-1})^{-1/2}(\hat{\theta}_i - \theta_{i,0}) \right| > Z_\alpha$$

où  $Z_\alpha$  est la valeur critique pour la valeur absolue d'une loi normale standard.

Compte tenu des hypothèses de l'article Jiang (1996), ce test n'est pas valide pour tester la nullité d'une composante. En pratique on peut vérifier que si on l'applique quand même et ce surtout pour les petits échantillons, il est peu puissant et conservatif (le niveau réel est nettement plus important que le niveau nominal).

Des calculs élémentaires quoique longs, montrent que dans le cas de la vraisemblance restreinte, l'information de Fisher vaut

$$I_\gamma = \left\{ \frac{1}{2} Tr(V_i V^{-1} Q V_j V^{-1} Q), i, j = 1, k \right\}$$

Un autre test possible qui est en général plus puissant quoique parfois non-conservatif est le **test du rapport de vraisemblance**. Si  $L_g$  est la log vraisemblance prise au maximum et  $L_p$  est la vraisemblance ou maximum sous l'hypothèse nulle, on montre que sous les mêmes hypothèses que précédemment

$$L_g - L_p \xrightarrow{\mathcal{L}} \frac{1}{2} \chi^2(k'),$$

où  $k'$  est la différence de dimension paramétrique entre les deux modèles : hypothèse générale, hypothèse nulle.

Ces deux tests : Wald et Rapport de Vraisemblance peuvent être indifféremment utilisés pour le maximum de vraisemblance comme pour le maximum de vraisemblance restreint.



## Chapitre 2

# Plans d'expériences randomisés

### 1 Introduction

Dans cette partie nous allons présenter l'idée forte suivante : la statistique n'a pas comme seul objet de "traiter des données" mais également d'en préparer le recueil pour en améliorer la qualité. D'importants gains sont possibles lors de cette étape. Les méthodes que nous allons présenter s'appliquent plutôt à des expériences de "labo" dans le sens le plus général possible, plutôt qu'à des situations où les données sont recueillies "comme elles viennent" par exemple dans les enquêtes.

Les deux buts de la planification sont : (1) de permettre une interprétation claire en évitant les **confusions**, (2) de maximiser la précision de l'expérience.

Pour illustrer le premier point, prenons l'exemple de la scolarisation en maternelle. Des études incontestables ont montré que les enfants scolarisés en maternelle ont de meilleurs résultats dans la suite de leur scolarité que les enfants qui ne rejoignent l'école qu'au primaire. Doit-on en déduire qu'il faut rendre la scolarisation en maternelle obligatoire pour lutter contre l'échec scolaire ? Une réponse directe : oui n'est pas possible. En effet deux interprétations sont possibles : (a) c'est effectivement la scolarité en maternelle qui améliore les résultats ; (b) dans la France actuelle les élèves qui ne vont pas en maternelle sont une minorité qui correspond à des groupes sociaux bien particuliers, ce qui peut expliquer la différence de réussite scolaire par l'origine sociale.

Dans cet exemple on pourra affiner l'analyse en contrôlant toute les variables indiquant la situation sociale, avec toujours le risque d'en oublier une. Mais il est clair que sans cette information complémentaire les données de départ sont **sans valeur** pour répondre à la question. Une solution de type planification à ce problème serait de définir un groupe d'enfants test et d'au hasard les répartir en deux groupes l'un qui serait scolarisé en maternelle l'autre non. Bien sûr, c'est impossible.

Pour illustrer le second point, considérons le problème de pesée de deux objets  $A$  et  $B$  avec une balance sans biais qui donne chaque résultat avec une erreur indépendante de variance  $\sigma^2$ . On suppose que la balance est capable de peser les deux objets ensemble.

**1ère méthode** : On pèse  $A$  et  $B$  séparément ; Le coût est de deux pesées la précision est  $\sigma^2$ .

**2ème méthode** : On pèse  $A+B$  et  $A-B$  ; Le coût est de deux pesées, la précision est  $\sigma^2/2$ , car les poids de  $A$  et  $B$  sont maintenant obtenues comme moyennes de deux pesées.

## 2 Nécessité de la randomisation

Comme nous allons essayer de vous en convaincre, la randomisation est la seule méthode qui évite les confusions (en fait elle en contrôle la probabilité), permet de faire une expérience équitable, de le prouver et enfin permet d'apprécier la précision des résultats.

Nous voulons comparer l'efficacité de deux médicaments A et B contre la grippe sur 40 malades.

expérience 1 On administre le médicament A aux 20 premiers malades qui se présentent. On note leur état d'amélioration, ensuite on administre B aux 20 suivants et on note leur état. A la fin de l'expérience, on calcule les moyennes et on déclare comme meilleur le médicament qui a la meilleure moyenne.

Cette expérience est désastreuse : (1) durant la durée de l'expérience la maladie, la température extérieure, la fatigue des personnes qui ont réalisé l'expérience ont pu évoluer : l'expérience n'est plus équitable dû à la confusion entre le temps et le médicament ; (2) Certains participants de l'expérience : malades ou médecins qui connaissent parfaitement les médicaments administrés peuvent fausser le résultat inconsciemment : c'est l'effet placebo bien connu.

expérience 2 Au fur et à mesure qu'un patient arrive en consultation, on alterne strictement A et B, le plan est donc ABABABABAB... Ce plan est certainement meilleur que le précédent mais souffre encore de deux gros défauts : en premier lieu la systématisme rend impossible le "double aveugle" le médecin et dans une certaine mesure le malade sauront toujours la nature du produit administré ; en second lieu on ne dispose pas de méthode statistique valide pour choisir entre les situations : A meilleur que B ; B meilleur que A ; A et B équivalents.

expérience 3 On pourrait construire une variante de l'expérience 2 où l'on essaierait de répartir au mieux les individus entre les deux groupes en fonction de l'âge, du poids, des antécédents. Ce plan séduisant a exactement les mêmes inconvénients que le précédent.

expérience 4 On tire au hasard 20 personnes parmi les 40 premiers malades auquel on administre A. Les autres reçoivent B. On pose un modèle conceptuel : on imagine avoir administré les deux médicaments à chacun des malades et note  $R_{ik}$  la réponse du malade  $i$  au médicament  $k$ . On pose

$$R_{ik} = m_i + a_k$$

qui suppose l'additivité entre l'effet du malade  $a_k$  et l'effet du médicament  $m_i$ . Pour se débarrasser d'une indétermination on pose  $\sum_k a_k = 0$ . Maintenant on note  $Y_{ij}$  la réponse du jème malade parmi ceux qui ont reçu le traitement  $i$ . Du modèle précédent on déduit

$$Y_{ij} = m_i + b_{ij}$$

où les  $b_{ij}$  sont tirés sans remise parmi les  $a_k$ . Les propriétés du tirage sans remise montrent que les  $b_{ij}$  sont d'espérance nulle et que leur matrice de variance (rangés par exemple en ordre lexicographique) vaut

$$\sigma^2(Id - J/n),$$

où  $n$  est le nombre total de données,  $J$  est la matrice  $n \times n$  composée de 1 et  $\sigma^2 = \frac{1}{n-1} \sum_{k=1, n} a_k^2$

Non seulement, le terme extra-diagonal  $\sigma^2/n$  est petit mais surtout il est constant : son influence sur les comparaisons entre traitements est nulle, de sorte que l'on peut faire une analyse de la variance sur les données présentes, ce qui permet de savoir si les médicaments sont significativement différents.

**Exercice 1** *Prouver l'affirmation ci-dessus :*

$$\text{Var}(Y_{i.} - Y_{i'}) = \frac{2\sigma^2}{r}$$

Enfin l'expérience 4 permet le travail en "double aveugle", ainsi elle garantit et prouve l'équité de l'expérience.

### 3 Plans d'expériences classiques

dans la présentation on utilisera les notations classiques suivantes :  $t, r, b, k, n$  sont respectivement les nombres de traitements, répétitions de chaque traitement, de blocs, d'unités par bloc et enfin de données.

#### Plan en randomisation totale

C'est le plan de l'expérience 4 précédente. Dans cas général, on a  $t$  traitements,  $rt$  unités expérimentales. On tire au hasard et sans remise,  $r$  unités pour le premier traitement, puis  $r$  pour le second etc... les données sont analysées par une analyse de la variance à 1 facteur : le traitement. A part dans les expérience médicales ce plan est en fait peu utilisé.

#### Plan en blocs complets

[Fisher 1931] Ce plan suit les trois principes de R. Fisher : répétition ; randomisation et contrôle local. Dans le même cadre que le plan en randomisation totale, on regroupe les  $rt$  unités expérimentales en  $r$  **blocs homogènes** de taille  $t$ . Dans un exemple médical les blocs peuvent par exemple correspondre aux sexe, à l'âge etc. Dans une expérience agronomique ce sera un ensemble de parcelles contiguës, dans toute expérience de labo, ce sera les unités traitées le même jour, par la même personne. Le principe du plan est le suivant : dans chaque bloc on alloue indépendamment une unité **exactement** à chaque traitement et ce de façon aléatoire.

*Exemple :* Une association de consommateurs désire comparer le fonctionnement de  $t$  aspirateurs. Les essais sont conduits par des ménagères membres de l'association. On suppose qu'il y a  $b$  ménagères et que chacune veut bien conduire  $t$  essais. Chaque ménagère a donc à sa disposition  $t$  "cases" pour un essai. L'ensemble de ces cases de taille  $tb$  constitue l'ensemble des unités, les ménagères constituent les blocs. Le plan revient à tirer au hasard l'ordre dans lequel chaque ménagère va expérimenter chaque aspirateur.

Pour être efficace le plan doit maximiser la variabilité inter-bloc et minimiser la variabilité intra-bloc : en d'autres termes il faut rendre les blocs le plus homogènes possible. Dans l'exemple, on aura intérêt à ce que les  $t$  essais d'une même ménagère soient aussi rapprochés que possible. Il est beaucoup moins gênant par contre qu'au total l'expérience se déroule sur un long laps de temps.

Des calculs analogues à ceux fait pour le plan en randomisation totale mais un peu plus complexes montrent que l'on peut valider une analyse de la variance à deux facteurs additifs : traitement et bloc. Cette analyse est équilibrée.

Le plan en blocs complets est quasi-toujours préférable au plan en randomisation totale. C'est le plan le plus employé. C'est celui que l'on essaiera d'utiliser a priori.

### Plan en blocs incomplets

**Plans équilibrés :** Supposons que nous ayons 9 bières à comparer (facteur d'intérêt : facteur traitement) et 12 dégustateurs (facteur bloc). Il est clair que passé un certain nombre de dégustations, un dégustateur est incapable de comparer ses sensations. On supposera donc qu'un dégustateur ne peut comparer que 3 bières. Nous avons donc 36 unités réparties en bloc de 3 et on propose la répartition suivante :

Deg. 1	1	2	3	Deg. 4	1	4	7
Deg. 2	4	5	6	Deg. 5	2	5	8
Deg. 3	7	8	9	Deg. 6	3	6	9
Deg. 7	1	5	9	Deg. 10	1	8	6
Deg. 8	4	8	3	Deg. 11	2	4	9
Deg. 9	7	2	6	Deg. 12	3	5	7

Cette répartition qui n'a rien d'aléatoire, pour l'instant, est équilibrée : chaque traitement se retrouve une fois et une seule exactement avec chaque autre traitement.

Un plan en blocs incomplets équilibré est un plan possédant la propriété ci-dessus, il est décrit (partiellement) par :

le nombre de traitements  $t$  (9 dans l'exemple)

le nombre de répétitions  $r$  (4)

le nombre de blocs  $b$  (12)

la longueur d'un bloc  $k$  (3)

l'indice de concurrence  $\lambda$  : le nombre de fois où deux traitements se retrouvent ensemble (1). En comptant de deux manières différentes le nombre de parcelles et le nombre de voisins, on obtient :

$$rt = bk$$

$$r(k - 1) = \lambda(t - 1)$$

Il n'existe pas des solutions pour toutes les tailles vérifiant ces équations. Il existe des tables de plans : Raghavarao (1971).

**Les lattices :** Nous présentons ci-dessous une méthode pour construire des plans qui sont sous certaines conditions équilibrés et qui conservent de toute façon de bonnes propriétés : les plans lattice.

**Définition 4 (lattices)** *Un  $(h, p^2)$  lattice est un plan pour  $p^2$  traitements avec  $p$  premier ou puissance de premier avec  $h$  répétitions et des blocs de longueur  $p$ . Aux  $p^2$  traitements sont associés deux facteurs  $A$  et  $B$  à  $p$  niveaux et dont les niveaux sont numérotés par des éléments du corps  $F_p$  à  $p$  éléments. Le plan est constitué de  $h$  **répliques** : ensembles de  $p$  blocs qui contiennent une fois et une seule chaque traitement. Dans la réplique 1 on confond*



$A$  avec les blocs, dans la réplique 2  $B$  et dans les autres successivement  $A + iB, i \in F_p \setminus \{0\}$ . On montre que le lattice  $(p+1), p^2$  est équilibré.

**Exercice 2** En utilisant la description de  $F_4$  comme l'ensemble  $\{0, 1, x, 1+x\}$  muni de l'addition modulo 2 et de la multiplication modulo 2 et le polynôme irréductible  $1+x+x^2$ , construire le lattice  $5, 4^2$ . Par exemple sur  $F_4$   $(1+x)(1+x) = 1+2x+x^2 = x$

**Définition 5 (Plans circulants)** Un plan circulant est un plan pour  $t$  traitements en  $t$  blocs de tailles  $t-1$ . Pour construire les blocs on élimine tour à tour chacun des traitements.

Un plan circulant est toujours équilibré, son indice de concurrence  $\lambda$  vaut  $t-2$ .

**Plans non équilibrés :** Quand il existe un plan équilibré, on montre qu'il est optimal (Pour plus de détail voir Druilhet, 19xx). malheureusement cela n'est pas toujours possible en particulier l'équilibre demande souvent un nombre de répétition élevé. Dans ce cas il existe des méthodes pour construire des plans conservant certaines propriétés, par exemple l'équilibre partiel (Coursol, 1980). mais de toutes façons les propriétés suivantes restent vraies **que le plan soit équilibré ou non.**

**Randomisation :** La randomisation se fait en deux étapes,

- (1) "mélange des blocs" : dans l'exemple on affecte un numéro de dégustateur à un dégustateur réel (M Dupond) au hasard ;
- (2) "mélange des traitements par bloc" : dans l'exemple, les trois bières devant être présentées à un dégustateur, le sont dans un ordre aléatoire.

**Analyse :** On montre que la randomisation valide un modèle avec des effets traitement fixes et des effets blocs aléatoires : c'est un modèle mixte. Si on ne dispose pas des moyens de traiter un tel modèle, on peut toutefois utiliser un modèle avec blocs et traitements fixes qui correspond à une légère perte d'information.

### Carrés latins

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 parasites. Si on veut équilibrer les relations du premier facteur avec chacun des deux autres, on peut être amené (avant randomisation) à utiliser la répartition suivante

peinture	orientation			
	N	S	E	O
A	1	2	3	4
B	2	3	4	1
C	3	4	1	2
D	4	1	2	3

On appelle *carré latin*, un plan pour trois facteurs au même nombre  $n$  de niveaux, comprenant  $n^2$  unités et tel que le nombre de répétitions d'une paire de niveaux pour deux facteurs est toujours 1.

**Randomisation :** La répartition du tableau ci dessus n'est pas très aléatoire, on satisfera à l'exigence de randomisation en faisant dans l'ordre que l'on voudra un "mélange des lignes" et "un mélange des colonnes". Cette randomisation valide, sous les mêmes hypothèses d'additivité que pour le plan en randomisation totale, une analyse de la variance à trois facteurs additifs.

La méthode de construction "en diagonale" fonctionne pour toute valeur de  $n$ . Mais il existe de nombreuses méthodes de construction. La plus classique est basée sur un groupe fini. On construit la valeur du troisième facteur comme la table d'addition de deux permutations du groupe (noté additivement).

**Plans en ligne et colonnes non équilibrés :** On peut s'intéresser à un plan dont les unités sont placées sur un réseau de  $l$  lignes et de  $c$  colonnes mais dont l'allocation des traitements n'a plus la belle propriété du carré latin. On utilise encore la même randomisation, mais elle valide maintenant une analyse avec un effet ligne et colonne aléatoire.

**Exercice 3** Vérifiez informatiquement que pour un carré latin on peut déclarer indifféremment les effets lignes et colonnes comme fixes ou aléatoires, cela ne change pas les résultats.

### Split-plot (parcelle subdivisée)

Exemple : on désire comparer 3 recettes de gâteau au chocolat et 6 températures de cuisson 180,(10)...,230. On réalise l'expérience suivante : chaque jour, on réalise 3 pâtes correspondant aux 3 recettes. Chaque pâte est divisée en 6 sous-partie cuites à chacune des 6 températures différentes. Après cuisson, on mesure le moelleux du gâteau en mesurant l'angle de rupture  $\alpha$  d'une tranche. On a donc pour expliquer cette variable quantitative trois facteurs : la température (que l'on décide de considérer comme qualitatif), la recette et le jour (facteur bloc). Ce qui est particulier à cet exemple est qu'il y a deux sources d'erreurs : une attachée à la mesure de l'angle et l'autre à la confection d'une pâte.

**Définition 6** On appelle plan split-plot un plan pour deux facteurs traitements. Le premier  $A$  à  $t$  niveaux et le second  $B$  à  $s$  niveaux. Pour le premier facteur, on construit un plan en blocs complets à  $rt$  "grandes unités" avec sa randomisation. Ensuite, chaque "grande unité" est divisée en  $s$  sous unités auxquelles sont affectées dans un ordre aléatoire les  $s$  valeurs du traitement  $B$

Si  $i, j, k$  sont les niveaux de  $A, B, bloc$  dans l'ordre, le modèle auquel conduit la randomisation est

$$Y_{ijk} = \mu + a_i + b_j + a * b_{ij} + bl_k + E_{ik} + \epsilon_{ijk} ; i = 1, t ; j = 1, s ; k = 1, r.$$

où  $E$  est l'erreur associée aux "grandes unités".

Ce modèle peut s'analyser bien évidemment à l'aide d'un programme de modèle mixte, mais il peut également s'analyser à partir des projections inter et intra "grande unité". La projection inter revient à travailler sur  $Y_{i.k}$  qui avec des contraintes classiques donne

$$Y_{i.k} = \mu + a_i + bl_k + E_{ik} + \epsilon_{i.k}.$$

Les deux aléas peuvent être confondus, on retrouve ainsi le modèle du plan en blocs complets. La projection intra est théoriquement basée sur les  $Y_{ijk} - Y_{i.k}$ . On montre qu'elle est équivalente au modèle complet dans lequel l'effet  $E_{jk}$  est supposé fixe, ceci à condition de se limiter à l'estimation et aux tests sur le facteur  $B$  et sur l'interaction  $A * B$ .

En conclusion on dit que le facteur  $A$  est totalement estimable inter-grandes unités et l'interaction  $A * B$  et le facteur  $B$  sont totalement estimable intra-grandes unités.



# Chapitre 3

## Plans fractionnaires

### 1 Introduction

On considère une réaction chimique qui dépend pour simplifier de trois facteurs : le  $pH$   $PH$ , avec une valeur standard de 7, la température,  $T$  avec comme valeur standard  $30^\circ C$ , et la dose  $D$ , avec comme valeur standard 100. On sait que l'on peut faire varier chacun de ces facteurs entre deux limites et on cherche à savoir s'il ont une influence sur la réponse, par exemple le rendement. On va comparer deux expériences

Expérience 1 : On fait varier d'abord le facteur  $PH$

On fait 4 répétitions à  $PH = 6.5$ ,  $T = 30^\circ$ ,  $D = 100$  et 4 répétitions à  $PH = 7.5$ ,  $T = 30^\circ$ ,  $D = 100$ .

On fait varier ensuite le facteur température

On fait 4 répétitions à  $PH = 7$ ,  $T = 25^\circ$ ,  $D = 100$  et 4 répétitions à  $PH = 7$ ,  $T = 35^\circ$ ,  $D = 100$ .

On fait varier enfin le facteur dose

On fait 4 répétitions à  $PH = 7$ ,  $T = 30^\circ$ ,  $D = 90$  et 4 répétitions à  $PH = 7$ ,  $T = 30^\circ$ ,  $D = 110$ .

Le coût total de cette expérience est de 24 unités et la puissance expérimentale est la comparaison de moyennes de 4.

Expérience 2 : On réalise toutes les 8 combinaisons entre les deux valeurs hautes et basses des trois facteurs. C'est à dire les huit unités

$(6.5, 25, 90)$ ,  $(6.5, 25, 110)$ ,  $(6.5, 35, 90)$ ,  $(6.5, 35, 110)$ ,

$(7.5, 25, 90)$ ,  $(7.5, 25, 110)$ ,  $(7.5, 35, 90)$ ,  $(7.5, 35, 110)$

Le coût total est maintenant de 8 unités. Si on fait l'analyse à l'aide d'un modèle additif à trois facteurs :  $H, T, D$ , ce modèle est orthogonal, on comparera donc des moyennes de 4 observations pour tester la significativité d'un facteur. Donc en première approximation, on a construit une expérience trois fois moins chère qui a la même puissance.

Remarquons bien, qu'en présence d'interactions entre les facteurs, il sera possible de les détecter dans la seconde expérience : en effet il reste 4 degrés de liberté dans la résiduelle, il est encore possible d'en occuper un ou deux pour un ou deux termes d'interaction. Par

contre comme l'expérience 1 est subdivisées en sous-expériences dans lesquelles on ne fait varier qu'un facteur à la fois, il ne sera pas possible de détecter la présence d'interaction.

En conclusion, quand un phénomène dépend de plusieurs facteurs, on a toujours intérêt à les étudier globalement. Le plan le plus simple est celui de l'expérience 2 où l'on fait toutes les combinaisons des niveaux (dans notre exemple chaque facteur a deux niveaux) de tous les facteurs. Il est appelé **plan factoriel complet**. Si il y a  $p$  facteurs tous à 2 niveaux, le plan factoriel complet demande  $2^p$  unités. Cela est souvent trop coûteux. On va donc faire une partie : **une fraction** du plan complet. Un tel plan est appelé **plan factoriel fractionnaire**

## 2 Cadre général pour des facteurs à deux niveaux

On considère  $p$  facteurs à deux niveaux noté  $-1, +1$ . On appelle traitement une combinaison  $(i_1, \dots, i_p)$  de ces facteurs. L'ensemble des traitements est de cardinal  $2^p$  et correspond au plan factoriel complet. Soit  $E$  l'espace des réponses aux divers traitements. Cet espace peut être vu comme l'ensemble des fonctions

$$[-1, +1]^p \rightarrow \mathbb{R}$$

ou comme l'espace vectoriel  $\mathbb{R}^{2^p}$ .

On se place dans le cas où  $n := 2^p$  est trop grand pour pouvoir expérimenter toutes les combinaisons : on va n'en faire qu'une partie, une fraction d'où le nom.

En premier, nous allons définir proprement les interactions multiples et les effets principaux. Dans  $E$  il existe des éléments particulier : les fonction coordonnées. On définit la  $k$ ième fonction coordonnée  $A_k$ , ( $1 \leq k \leq p$ ) par

$$(i_1, \dots, i_p) \xrightarrow{A_k} i_k$$

Soit  $B \subset \{1, \dots, p\}$ , on définit la fonction  $A^B$  élément de  $E$  par

$$A^B := A_1^{\epsilon_1} \dots A_p^{\epsilon_p} \quad ; \quad \epsilon_k = 1 \text{ si } k \in B \quad ; \quad \epsilon_k = 0 \text{ sinon}$$

**Proposition 3** *Quand  $B$  varie dans l'ensemble des parties de  $\{1, \dots, p\}$ , les  $A^B$  forment une base orthogonale de  $E$  de norme  $\sqrt{n}$ .*

On définit les espaces suivants pour  $B$  partie de  $\{1, \dots, p\}$ .

$V_B$  =fonction des seules coordonnées dans  $B$

$W_B$  orthogonal dans  $V_B$  de tous les  $V_D$  pour  $D \subset B, D \neq B$

**Exemple 1** *Par exemple si  $p = 2$  un élément de  $E$  peut s'écrire comme un tableau 2 de la forme*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

*Un élément de  $V_{\{1\}}$  est de la forme*

$$\begin{pmatrix} a & a \\ b & b \end{pmatrix}.$$

Un élément de  $W_{\{2\}}$  :

$$\begin{pmatrix} a & -a \\ a & -a \end{pmatrix}.$$

**Proposition 4**  $\dim V_B = 2^{|B|}$ , où  $|B|$  est le nombre d'éléments de  $B$   
si  $D \subset B$ ,  $A^D \in V_B$ ,  
 $\{A^D \text{ pour } D \subset B\}$  engendrent  $V_B$ ,  
 $A^B \in W_B$ ,  
 $A^B$  engendrent  $W_B$  qui est donc de dimension 1.

$W_B$  est défini comme l'espace de l'interaction entre les facteurs de  $B$  avec les exceptions suivantes : si  $B$  est réduit à un facteur, il s'agit alors de l'effet principal de ce facteur, si  $B = \emptyset$ ,  $W_\emptyset$  est l'espace de la "moyenne générale". Il est de dimension 1 quelle que soit la taille de  $B$ .

Les différents résultats de la proposition s'enchaînent aisément de sorte que leur démonstration est aisée.

**Définition de la valeur d'un effet principal ou d'une interaction :** La vraie réponse  $f$  au traitement est un élément de  $E$  qui se décompose sur la base des générateurs  $A^B$ . Nous appellerons "effet de  $A^B$ " noté  $e(A^B)$  le coefficient de cette décomposition.

$$f = \sum_{B \subset \{1, \dots, p\}} e(A^B) A^B.$$

**Exercice 4** construisez les différents espaces :  $V_\emptyset, V_{\{1\}}, V_{\{2\}}, V_{\{1,2\}}, W_\emptyset, W_{\{1\}}, W_{\{2\}}, W_{\{1,2\}}$  dans le cas de  $p = 3$  facteurs.

### Exemple avec trois facteurs.

Le tableau suivant décrit le plan factoriel complet ainsi que les différents générateurs avec nos notations

$A^\emptyset$	$A^{\{1\}}$	$A^{\{2\}}$	$A^{\{3\}}$	$A^{\{1,2\}}$	$A^{\{1,3\}}$	$A^{\{2,3\}}$	$A^{\{1,2,3\}}$	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)

Notons  $A := A^{\{1\}}$ ,  $B := A^{\{2\}}$ ,  $C := A^{\{3\}}$ . Choisissons les 4 données pour lesquelles  $ABC = 1$ , ce qui définit un "demi plan" ou "fraction". Même en supposant observer la réponse sans erreur, on est amené à résoudre le système de quatre équations :

$$f(-1, -1, +1) = e(1) + e(ABC) - e(A) - e(BC) - e(B) - e(AC) + e(C) + e(AB)$$

$$f(-1, +1, -1) = e(1) + e(ABC) - e(A) - e(BC) + e(B) + e(AC) - e(C) - e(AB)$$

$$f(+1, -1, -1) = e(1) + e(ABC) + e(A) + e(BC) - e(B) - e(AC) - e(C) - e(AB)$$

$$f(+1, +1, +1) = e(1) + e(ABC) + e(A) + e(BC) + e(B) + e(AC) + e(C) + e(AB)$$

D'après les relations d'orthogonalité déjà prouvées, ce système a une solution unique en les inconnues

$$e(1) + e(ABC)$$

$$e(A) + e(BC)$$

$$e(B) + e(AC)$$

$$e(C) + e(AB).$$

Dans le tableau restreint correspondant à  $ABC = 1$ , certaines relations sont vérifiées entre les vecteurs. Elle découlent directement de la relation qui a défini la fraction :  $1 = ABC$ ,  $A = BC$ ,  $B = AC$ ,  $C = AB$ . Elles impliquent directement les confusions observées. En effet dans l'écriture générale :

$$f = \sum_{B \subset \{1, \dots, p\}} e(A^B) A^B,$$

certaines  $A^B$  sont confondus sur la fraction, on regroupe donc les coefficients.

### 3 Méthode des facteurs de base

On appellera plan  $2^{p-q}$  ( $p > q$ ), un plan comprenant  $2^{p-q}$  unités pour  $p$  facteurs à deux niveaux. Le taux de fraction est donc  $2^{-q}$ . Une méthode de construction qui donne des *fractions régulières* (terme que l'on définira ultérieurement) est la méthode des facteurs de base : on construit le plan factoriel complet pour les  $p - q$  premiers facteurs. On définit ensuite les valeurs des autres facteurs en fonction des  $p$  premiers (sous forme de produit exclusivement). On appelle *relations de définition* ces  $q$  relations que l'on exprimera sous la forme canonique  $1 = \dots$ . Il est facile de vérifier que la fraction ainsi définie est exactement le sous ensemble du plan complet qui vérifie les relations de définition. On définit également *les relations complètes de définition* qui sont les  $2^q$  terme égaux que l'on obtient en combinant les relations de définition. Ces relations donnent *l'alias* de 1 c'est à dire les termes confondus avec 1. A partir de cet alias et par combinaison, on obtient l'alias d'un terme quelconque donné qui comprend  $2^q - 1$  termes.

**Exemple 2** *Considérons le plan  $2^{5-2}$  pour 5 facteurs  $A, B, C, D, E$  avec  $A, B, C$  comme facteurs de base et les relations de définition :  $D = AB$ ,  $E = AC$ . Il s'agit donc d'un quart de plan à 8 unités. Les relations complètes de définition sont*

$$1 = ADB = ACE = DCBE.$$

*En effet  $A^2 = B^2 = \dots = E^2 = 1$ . Il y a bien 4 effets confondus. L'alias de  $A$  par exemple comprend trois termes :*

$$A = DB = CE = ABCDE.$$

*Il y a donc confusion dans cette fraction entre l'effet principal du facteur  $A$  et les interactions  $D * B$  et  $C * E$ .*



### Estimation des effets

On suppose que les observations sont faite avec une variance  $\sigma^2$  et qu'elle sont non corrélées. Choisissons une suite de générateurs  $:A^{B_1}, \dots, A^{B_m}, m \leq 2^{p-q}$  telle qu'il y ait au plus un représentant de chaque alias. On n'est pas obligé de les prendre tous. Alors le modèle linéaire

$$Y(i_1, \dots, i_p) = \sum_{i=1}^m \tilde{e}(A^{B_i}) A^{B_i} + \epsilon_{i_1, \dots, i_p},$$

où  $\tilde{e}(A^{B_i})$  est l'alias de  $e(A^{B_i})$  au sens de la somme de effets de tous les générateurs qui lui sont confondus, est orthogonal sur l'ensemble des données du plan fractionnaire. La matrice d'information :  $X'X$  vaut  $nId$  ( $n$  est le nombre de données :  $2^{p-q}$ ) et les estimateurs

$$\hat{\tilde{e}}(A^{B_i}) = \frac{\langle Y, A^{B_i} \rangle}{n},$$

sont non corrélés de variance  $\sigma^2/n$ .

Un fraction qui a la propriété que deux effets sont orthogonaux ou confondus est appelé **une fraction régulière**.

## 4 Plan pour l'étude des effets principaux et des interaction doubles

Un plan est dit de résolution  $\rho$  entière si

$$\rho = \inf\{\text{nb de symboles des éléments de l'alias de 1}\}$$

Exemples :

Résolution III : tous les effets principaux sont non confondus.

Résolution IV : un effet principal ne peut être confondu avec un interaction, mais deux interactions peuvent être confondues.

Résolution V : on peut poser un modèle avec toutes les interactions et effets principaux sans confusion.

De manière générale, la résolution V est considérée comme suffisante dans toutes les situations, alors que la résolution III est considérée comme une propriété minimale.

**Définition 7 (Aberration)** *Cette notion est un raffinement de la notion de résolution. L'aberration d'un plan pour  $p$  facteurs est un  $p$ -uple dont l'élément  $i$  contient le nombre de symboles de  $i$  lettres dans l'alias de 1 (1 exclu). Par exemple l'aberration du plan de la partie 3 vaut :  $(0, 0, 2, 1, 0)$*

Les aberrations sont munies de l'ordre lexicographique, on parle de plan de minimum d'aberration.

**Proposition 5** *Si le plan est de résolution III, tous les effets principaux sont estimables donc en comptant les degrés de liberté, le nombre  $p$  de facteurs vérifie*

$$p \leq n - 1$$

Ce maximum est atteint dans le sens où pour tout  $r$  il existe un plan de résolution III avec  $2^r$  unités et  $2^r - 1$  facteurs.

Une fraction régulière de résolution IV avec  $2^r$  unités comprend au maximum  $2^{r-1}$  facteurs. Ce maximum est atteint.

En résolution V, il n'y a pas de résultat général, on connaît simplement le nombre de facteurs maximaux pour les petites valeurs

$r$	4	5	6	7	8	9
nombres d'unités = $2^r$	16	32	64	128	256	512
nb max de facteurs $p$	5	6	8	11	17	23
ddl du modèle	16	22	37	67	154	277
ddl du modèle avec 1 Fac. de plus	22	29	46	79	172	301

### Compléments

- Quand le plan est de taille trop importante pour pouvoir être conduit de manière homogène, il doit être découpé en blocs. Les facteurs blocs peuvent être considérés comme des facteurs ordinaires sauf que les notions de résolution et aberration n'ont plus la même pertinence. Le mieux est de regarder en détail les confusions en étant bien conscient que tout effet confondu avec un facteur bloc sera totalement non-estimable intra-bloc (voir exemple ci-dessous).
- Si un ou plusieurs des facteurs ont un nombre de niveaux égaux à 4, 8 (ou plus généralement une puissance de 2) on peut se ramener au cas précédent en recodant les niveaux à l'aide de 2 ou 3 pseudo-facteurs : Si  $A$  possède 4 niveaux notés 1, 2, 3, 4 on peut le recoder à l'aide de 2 pseudo-facteurs en utilisant la table suivante :

$A$	$A_1$	$A_2$
1	1	1
2	1	-1
3	-1	1
4	-1	-1

- Exemple. On veut expérimenter 5 facteurs :  $A, B, C, D, E$  à l'aide de 32 unités ce qui correspond à un plan complet. Malheureusement on considère qu'il est impossible de réaliser plus de 8 unités de manière homogène de sorte qu'il faut introduire un facteur bloc  $BL$  à 4 niveaux qui sera codé par deux pseudo-facteurs  $B_1$  et  $B_2$ . On est donc amené à chercher un plan  $2^{7-2}$ . Celui donné par la table ci-dessous  $B_1 = ABCD$   $B_2 = ABDE$  amène à confondre l'interaction  $CE$  avec  $B_1B_2$  qui est un effet bloc. On laisse à titre d'exercice, le soin au lecteur de vérifier que le plan suivant est meilleur  $B_1 = ABC$   $B_2 = CDE$ .
- Plus de deux niveaux. Les plans fractionnaires utilisent dans le cas général des techniques de corps finis qui sont basés sur des nombres premiers. Pour ces raisons il n'est possible que de travailler que sur un seul nombre premier : on peut donc faire des plans pour des facteurs à 2, 4 ou 8 niveaux comme pour des facteurs à 3, 9 ou 27 niveaux, mais il est impossible de mélanger les deux. Pour des facteurs à 3 niveaux il faut travailler non plus avec  $\{-1, 1\}$  mais avec les racines cubiques de l'unité  $\{1, j, j^2\}$ , ce qui amène à une présentation plus technique.

**Exercice 5** Choisir un plan permettant d'expérimenter 7 facteurs  $A - G$  en estimant les effets principaux en présence d'éventuelles interactions. Donner les relations complètes de

définitions et l'alias du facteur A.

**Exercice 6** On veut construire un plan pour 5 facteurs A – E, on veut pouvoir estimer sans confusion les effets principaux et les interactions suivantes :  $A*D$ ,  $A*E$ ,  $B*D$ ,  $B*E$ .

Proposez une solution de dimension minimale.

**Exercice 7** Combien faut-il d'unités au moins pour expérimenter 12 facteurs en résolution IV et V respectivement ?

**Exercice 8 (repliement de plans)** Un Expérimentateur réalise le plan  $2^{5-2}$  définie par les relations  $D = ABC$  et  $E = BC$ . Plus tard il réalise un plan identique sauf que tous les signes ont été inversés.

- i. Quelle est la résolution du plan ainsi défini ?
- ii. Aurait on choisi ce plan si, dès le départ, on avait su disposer de 16 unités ?

## 5 Méthode Tagushi

Certains facteurs d'une expérience que l'on peut contrôler en laboratoire peuvent ne pas être contrôlés en utilisation normale. Un des apports de Tagushi est d'avoir proposé des plans pour étudier l'influence de ces facteurs sur la variabilité du résultat.

**Exemple 3** on veut régler un fraise pour une certaine performance donnée. Mais dans l'utilisation future de cette fraise il y aura certains facteurs incontrôlés : la température de l'atelier, le degré d'usure de la fraise, la température de l'huile de refroidissement.

On construit ainsi un plan fractionnaire pour les facteurs contrôlés en utilisation normale et un autre plan fractionnaire pour les facteurs non contrôlés en utilisation normale. On "croise" ensuite les plans. Si le dispositif doit être réglé sur une spécification précise. On va rechercher le réglage de facteurs contrôlés en utilisation normale qui minimise l'Erreur Quadratique Moyenne (EQM), c'est à dire le carré du biais plus la variance où ce biais et cette variance se calculent sous la loi de probabilité donnés par le plan sur les facteurs non contrôlés en utilisation normale.

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=AB	/	/	/	/	/	/	/	/
8	Plan Complet	IV $2^{4-1}$  D=ABC	III $2^{5-2}$  D=AB E=AC	III $2^{6-3}$  D=AB E= AC F=BC	III $2^{7-4}$  D=AB E=AC F=BC G=ABC	/	/	/	/
16	Plan Complet		V $2^{5-1}$  E=ABCD	IV $2^{6-2}$  E=ABC F=BCD	IV $2^{7-3}$  E=ABC F = BCD G = ACD	IV $2^{8-4}$  E=BCD F=ACD G=ABC H=ABD	III $2^{9-5}$  E=ABC F=BCD G=ACD H=ABD I=ABCD	III $2^{10-6}$  E=ABC F=BCD G=ACD H=ABD I=ABCD J=AB	III $2^{11-7}$  E=ABC F=BCD G=ACD H=ABD I=ABCD J=AB K = AC
32			Plan Complet	VI $2^{6-1}$  F=AB CDE	IV $2^{7-2}$  F=ABCD G=ABDE	IV $2^{8-3}$  F=ABC G=ABD H=BCDE	IV $2^{9-4}$  F=BCDE G=ACDE H=ABDE I=ABCE	IV $2^{10-5}$  F=ABCD G=ACDE H=ABDE I=ACDE J=BCDE	IV $2^{11-6}$  F=ABC G=BCD H=CDE I=ACD J=ADE K =BDE
64				Plan Complet	VII $2^{7-1}$  G=ABC DEF	V $2^{8-2}$  G=ABCD H=ABEF	IV $2^{9-3}$  G=ABCD H=ACEF I=CDEF	IV $2^{10-4}$  G=BCDF H=ACDE I=ABDE J=ABCE	IV $2^{11-5}$  G=CDE H=ABCD I=ABF J=BDEF K=ADEF
128					Plan Complet	VIII $2^{8-1}$  H=ABC DEFG	VI $2^{9-2}$  H=ACD FG I=BCEFG	V $2^{10-3}$  H=ABCG I=BCDE J=ACDF	V $2^{11-4}$  H=ABCG I=BCDE J=ACDF K=ABC DEFG

TAB. 3.1 – Tableau de plans fractionnaires de minimum d'aberration. En ligne, le nombre d'unités  $2^{p-q}$ , en colonne le nombre de facteurs, une solution est donnée lorsqu'il existe un plan de résolution III au moins. Chaque cellule indique de haut en bas : la résolution, le nom du plan, les relations de définition.

# Chapitre 4

## Surfaces de Réponses, plans isovariants

### 1 Cadre de l'étude

On considère une variable  $Y$  éventuellement influencée par  $m$  variables quantitatives  $X_1, \dots, X_m$ . On note  $x := (x_1, \dots, x_m)$  une réalisation possible de l'ensemble des variables explicatives. On suppose :

- $E(Y)$  au point  $x$  est une fonction polynomiale de degré  $q$  de  $x$
- Toutes les observations de  $Y$  sont non-corrélées et de même variance (même si on répète plusieurs fois le même  $x$ ).

En ordonnant arbitrairement les unités du plan, on obtient un modèle linéaire classique non-gaussien.

$$E(Y) = X\beta \quad ; \quad \text{Var}(Y) = \sigma^2 Id. \quad (4.1)$$

On note  $p$  la dimension de  $\beta$ . Certaines colonnes de la matrice  $X$  sont liés fonctionnellement : certaines sont des produits ou puissances d'autres. Nous supposons qu'elle ne sont pas liés linéairement de sorte que le modèle (4.1) est régulier. On note  $X(x)$  la ligne de  $X$  qui correspondrait à l'observation au point  $x$  (Ce point ne fait pas forcément partie des unités du plan correspondant au modèle (4.1)).

**Définition 8** *On dit que le plan est isovariant (rotatable en anglais) si  $\text{Var}(X(x)\hat{\beta})$  ne dépend de  $x$  qu'à travers  $|x|$  où  $\hat{\beta}$  est l'estimateur classique des moindres carrés.*

### 2 Conditions d'isovariance

Pour examiner les conséquences de la définition nous introduisons les notations suivantes.

-Si  $P$  est une transformation orthogonale de  $\mathbb{R}^m$  il existe un endomorphisme (une matrice si on préfère) unique  $Q_P$  tel que

$$Q_P(X'(x)) = \left( X(P(x)) \right)'. \quad (4.2)$$

Comme  $P' = P^{-1}$ , il est facile d'en déduire que  $Q_{P'} = (Q_P)^{-1}$ .

**Lemme 1** Soit  $A$  une matrice symétrique définie positive de taille  $p, p$ . Alors il y a équivalence entre

- $Tr(A) = Tr(A^{-1}) = p$

et

- $A = Id$

Nous laissons la démonstration de ce lemme à titre d'exercice.

Nous notons  $M$  la matrice des moments, qui n'est rien d'autre que la matrice d'information (qui sera elle notée  $N$ ) divisée par le nombre de données :

$$M = \frac{1}{n} X'X = \frac{1}{n} N$$

Les éléments de  $M$  sont du type

$$[\delta_1, \dots, \delta_m] := \frac{1}{n} \sum_{i=1}^n X_{i1}^{\delta_1}, \dots, X_{im}^{\delta_m},$$

où  $X_{ji}$  est l'observation de la variable  $j$  sur l'unité  $i$ . La matrice  $M$  contient les éléments ci-dessus pour  $\delta_1 + \dots + \delta_m \leq 2q$

**Proposition 6** Le plan est isovariant si et seulement si  $M$  est isovariante c'est à dire que pour toute transformation orthogonale  $P : M = Q_P M Q_P'$

*Démonstration :* Condition suffisante : supposons  $M$  isovariante. Pour simplifier on note  $X(x)$  écrit en vecteur ligne par :  $\mathbf{X}$  et  $Q_P : Q$ . On a alors

$$X(P(x)) = \mathbf{X}Q',$$

et également

$$\frac{1}{\sigma^2} \text{Var}((P(x))\hat{\beta}) = \mathbf{X}Q' \frac{1}{\sigma^2} \text{Var}(\hat{\beta}) Q \mathbf{X}' = \mathbf{X}Q' (X'X)^{-1} Q \mathbf{X}' = \mathbf{X} (X'X)^{-1} \mathbf{X}'.$$

Passons maintenant à la condition suffisante : nous supposons donc maintenant que le plan est isovariant. Remarquons d'abord que

$$Tr(\text{Var}(X\hat{\beta})) = \sigma^2 Tr(X(X'X)^{-1}X') = \sigma^2 Tr((X'X)^{-1}X'X) = \sigma^2 p.$$

Soit maintenant le plan transformé par  $P$ . L'équation 4.2 montre que la matrice de ce plan vaut  $X(Q_P)'$ . Comme le plan est isovariant (vérifier ce dernier point en détail)

$$\text{Var}(X(Q_P)'\hat{\beta}) = \text{Var}(X\hat{\beta}).$$

Donc

$$\sigma^2 p = Tr(\text{Var}(X\hat{\beta})) = Tr(\text{Var}(X(Q_P)'\hat{\beta})) = \sigma^2 Tr(X(Q_P)'I^{-1}(Q_P)X')$$

et

$$p = Tr((Q_P)'I^{-1}(Q_P)I) \tag{4.3}$$

On obtient exactement le même résultat en remplaçant  $I$  par  $M$ . On pose donc

$$A := (M^{1/2}Q'_P M^{-1}Q_P M^{1/2})$$

La relation (4.3) implique que  $Tr(A) = p$ . Il est facile de vérifier par ailleurs que

$$A^{-1} = (M^{-1/2}(Q_P)^{-1}M(Q'_P)^{-1}M^{-1/2}).$$

Donc  $Tr(A^{-1}) = Tr((Q'_P)^{-1}M^{-1}(Q_P)^{-1}M^{-1}) = p$ , car  $(Q_P)^{-1} = Q_{P'}$  et il suffit d'appliquer le lemme.  $\square$ .

Nous en venons maintenant au résultat principal de cette section.

**Théorème 2 (Box et Hunter 1957)** *Avec nos notations, la matrice des moments est isovariante si et seulement si tout moment d'ordre  $\delta = \sum_{j=1}^m \delta_j \leq 2q$*

(i) *est nul si l'un des  $\delta_j$  est impair*

(ii) *est égal à*

$$\frac{\prod_j \delta_j!}{2^{\delta/2} \prod_j (\delta_j/2)!} \mu_\delta$$

*sinon.*

*Les  $\mu_\delta$  sont certaines constantes.*

L'interprétation de ce théorème peut se faire sur l'exemple incontournable des modèles quadratiques qui correspondent à  $q = 2$ . Dans ce cas on doit considérer les moments jusqu'à l'ordre 4. Compte tenu de (i), les moments non nuls ne peuvent être que ceux qui correspondent à

- Un  $\delta_j = 2$ , les autres  $\delta_{j'}$  nuls. Il doivent être tous égaux et le théorème nous dit que l'on note  $\mu_2$  leur valeur commune.

- Deux  $\delta_j$  égaux à 2, les autres nuls. Le théorème nous dit que l'on note maintenant  $\mu_4$  leur valeur commune .

- Un  $\delta_j = 4$ , les autres  $\delta_{j'}$  nuls. Il doivent être tous égaux et leur valeur est maintenant  $3\mu_4$  .

Nous reviendrons sur cet exemple ultérieurement. *Démonstration du théorème :*

On pose

$$F(t) := \frac{1}{n} \sum_i (1 + x_{i1}t_1 + \dots + x_{im}t_m)^{2q}$$

Un petit peu de combinatoire montre que le monôme d'ordre  $(\delta_1, \dots, \delta_m)$  a pour coefficient

$$\frac{(2q)!}{(2q - \delta)! \prod_j \delta_j!} [\delta_1, \dots, \delta_m]$$

Supposons la matrice des moments isovariante, alors les  $[\delta_1, \dots, \delta_m]$  le sont également et cela implique que pour toute transformation orthogonale  $P$  :

$$F(P(t)) = F(t)$$

$F$  est en fait une fonction de la valeur absolue de  $t$  et comme c'est un polynôme

$$F(t) = \sum_{\delta=0,q} a_\delta (\|t\|^2)^\delta = \sum_{\delta=0,q} a_\delta \sum' \frac{\delta!}{\prod_{j=1,m} \delta_j!} t_1^{2\delta_1} \dots t_m^{2\delta_m}$$

où la somme  $\sum'$  porte sur les  $\delta_j \leq 0$  tels que  $\delta_1 + \dots + \delta_m = \delta$ . En identifiant on obtient le résultat.  $\square$ .

### 3 Plans composites centrés de Box et Wilson

Nous allons nous concentrer notre attention sur les modèles quadratiques ( $q = 2$ ) pour  $m$  facteurs. Un plan composite est constitué de trois parties :

- une partie factorielle composée du plan factoriel complet  $\{-1, +1\}^m$  (on peut montrer qu'en fait une fraction régulière de résolution V suffit).
- d'une partie axiale composée de  $2m$  points à distance  $\alpha$  de l'origine et situés sur les axes ( $\alpha$  reste à déterminer).
- la partie centrale composée de  $n_0$  points au centre 0 ( $n_0$  reste à déterminer).

La distribution du plan est invariante

- par toute symétrie par rapport à une hyperplan d'équation  $x_j = 0$ .
- Ainsi que par toute permutation des coordonnées.

Deux conséquences immédiates sont que  $[\delta_1, \dots, \delta_m]$  est nul dès qu'un  $\delta_j$  est impair et que  $[\delta_1, \dots, \delta_m] = [\delta_{\sigma(1)}, \dots, \delta_{\sigma(m)}]$

Soient  $F$  le nombre de points de la partie factorielle du plan, il est immédiat que la matrice d'information  $N$  vérifie (en notant dans l'ordre : la constante, les effets linéaires, les carrés et enfin les produits

$$M = \begin{bmatrix} n & 0 & \dots & .0 & b & \dots & .b & 0 & \dots & .0 \\ 0 & b & 0\dots & .0 & 0 & \dots & .0 & 0 & \dots & .0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ b & 0 & \dots & .0 & c & d\dots & .d & 0 & \dots & .0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ b & 0 & 0\dots & .0 & d & \dots d & c & 0 & \dots & .0 \\ 0 & 0 & \dots & .0 & 0 & \dots & .0 & a & 0\dots & .0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & .0 & 0 & \dots & .0 & 0 & .0 & .a \end{bmatrix} \quad (4.4)$$

avec

$$b := \sum_{i=1,n} x_{i,1}^2 = F + 2\alpha^2 \quad ; \quad d := a := \sum_{i=1,n} x_{i,1}x_{i,2} = F \quad ; \quad c := \sum_{i=1,n} x_{i,1}^4 = F + 2\alpha^4$$

le plan est donc isovariant si  $\alpha = F^{1/4}$

Remarquez que pour  $m = 2$  et  $m = 4$  les points de la partie axiale et factorielle sont situés sur la même sphère. Pour les autres valeurs de  $m$  les points ont seulement qu'approximativement la même norme.

Nombre de variables $m$	2	3	4	5	5	6	6
Taille p. fact. $F$	4	8	16	32	16	64	32
taille p. axiale	4	6	8	10	10	12	12
Point centraux							
isovariance	$\leq 1$	$\leq 1$	$\leq 1$	$\leq 1$	$\leq 1$	$\leq 1$	$\leq 1$
précision uniforme	5	6	7	10	6	15	9
orthogonalité	8	12	12	17	10	24	15
$\alpha$	1.41	1.68	2	2.38	2	2.83	2.38

**Tableau 2 :** Récapitulatif des paramètres et des tailles des plans composites centrés.



## Chapitre 5

# Bibliographie

- Box, G., Hunter, W.G., Hunter J.S. (1978). *Statistics for experimenters*. Wiley, New-York.
- Coursol, J. (1980). *Technique statistique des modèles linéaires*. Cimpa, Nice.
- Dacunha-Castelle, D. & Duflo, M. (1982, 1983). *Probabilités et Statistiques , 1 Problèmes à temps fixe , 2. Problèmes à temps mobiles*. Masson, Paris.
- Droesbeke, J-J., Fine J. & Saporta G. (1997) (Ed.). *Plans d'expériences, Application à l'entreprise*. Technip.
- Jiang, J. (1996). REML estimation : Asymptotic behaviour and related topics. *Annals of statistics*, 24, 255-286.
- Raghavarao , D. (1971). *Construction and combinatorial problems in design of experiments*. Wiley, New-York.



# Table des matières

<b>1</b>	<b>Modèles mixtes</b>	<b>3</b>
1	Analyse de la variance multivariante . . . . .	3
2	Modèles à effets aléatoires et mixtes . . . . .	5
2.1	Facteurs croisés et hiérarchisés . . . . .	5
2.2	Modèles mixtes équirépétés . . . . .	6
2.3	Modèles mixtes généraux . . . . .	7
2.4	Estimation des effets fixes . . . . .	8
2.5	Estimation par MIVQUE dans un modèle mixte . . . . .	8
2.6	Estimation par maximum de vraisemblance restreinte . . . . .	9
2.7	Tests . . . . .	10
<b>2</b>	<b>Plans d'expériences randomisés</b>	<b>13</b>
1	Introduction . . . . .	13
2	Nécessité de la randomisation . . . . .	14
3	Plans d'expériences classiques . . . . .	15
<b>3</b>	<b>Plans fractionnaires</b>	<b>21</b>
1	Introduction . . . . .	21
2	Cadre général pour des facteurs à deux niveaux . . . . .	22
3	Méthode des facteurs de base . . . . .	24
4	Plan pour l'étude des effets principaux et des interaction doubles . . . . .	25
5	Méthode Tagushi . . . . .	27
<b>4</b>	<b>Surfaces de Réponses, plans isovariants</b>	<b>29</b>
1	Cadre de l'étude . . . . .	29
2	Conditions d'isovariance . . . . .	29
3	Plans composites centrés de Box et Wilson . . . . .	32
<b>5</b>	<b>Bibliographie</b>	<b>33</b>