

REVUE DE STATISTIQUE APPLIQUÉE

V. COHEN

Introduction aux plans d'expérience

Revue de statistique appliquée, tome 37, n° 2 (1989), p. 17-45

http://www.numdam.org/item?id=RSA_1989__37_2_17_0

© Société française de statistique, 1989, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

INTRODUCTION AUX PLANS D'EXPERIENCE

V. COHEN

RESUME

Après avoir indiqué brièvement pourquoi et comment interviennent les plans d'expérience (en particulier, en pilotage de la qualité), nous mentionnons quelques problèmes liés à leur construction et à leur usage, en soulignant quelques spécificités liées à leur utilisation actuelle. Un cas d'espèce, le plan factoriel 2^n , est présenté de manière un peu plus détaillée, à titre d'illustration.

Mots-clés : Plan d'expérience, fonction de réponse, plan factoriel, fractionnement, plan réduit, qualité.

Plan

Introduction

1. Cadre de l'étude

1.1 Problèmes généraux

1.2 Problèmes spécifiques

2. Etude d'un cas : le plan factoriel 2^n

2.1 Exemples : plans 2^2 et 2^3

2.2 Représentation matricielle

2.3 Modèle probabiliste sous-jacent

2.4 Fractionnement en blocs

2.5 Plans réduits

2.6 Généralisation et problèmes connexes

Conclusion

Le plan d'expérience constitue, selon le mot de R. FISHER, "un essai technologique d'utilisation maximale (et, ajouterons-nous, de collecte efficace et économique) des données". Depuis son invention, vers 1935, [10], le domaine a connu un développement considérable, surtout dans les 30 années qui ont suivi.

Il est remarquable qu'issu de préoccupations extrêmement concrètes de l'agronome ou de l'industriel, il soit assez vite apparu profondément lié à des champs mathématiques réputés abstraits tels que la géométrie finie et les corps de Galois [1,11] (et qu'il y ait suscité des recherches), mais qu'en même temps, soient apparues des connexions avec d'autres branches des mathématiques appliquées comme la théorie de l'information [2].

La présentation qui suit ne constitue qu'une introduction à un domaine couvert depuis quelques dizaines d'années par beaucoup d'excellents ouvrages [7,8,9,14,17]. Les concepts et méthodes qu'il met en oeuvre trouvent aujourd'hui une nouvelle actualité d'une part en raison de l'importance prise par la *transmission et la protection de l'information*, d'autre part à cause de la primauté de la *qualité industrielle en environnement* à la fois très diversifié et très concurrentiel.

Je ne saurais oublier les exposés de M. VESSEREAU dont la seconde partie de cet article est largement tributaire sans en atteindre, sans doute, la clarté.

La gestion fonctionnelle de la qualité d'un produit conduit en effet :

dans une première phase, à rechercher les déterminants de cette qualité, c'est-à-dire à mettre en évidence les liens entre :

- un certain nombre de variables liées à la conception (par exemple, type de matériau utilisé), ou à son utilisation (par exemple, conditions climatiques d'emploi),
- un ou plusieurs indicateurs de qualité : performance, durée de vie, maintenabilité, etc ...

dans une seconde phase, à rechercher les configurations réalisables procurant les meilleurs résultats (en un sens à préciser).

La première phase est plutôt cognitive puisqu'elle vise à obtenir une meilleure connaissance du comportement du produit dans différentes situations, la seconde phase, plutôt normative puisqu'elle tend à optimiser ou du moins à améliorer le produit selon des critères choisis.

Mais la première phase comportera certains aspects normatifs dans la mesure où le choix des variables candidates à expliquer la qualité préjuge quelque peu des actions que l'on est en mesure de mener pour l'améliorer.

Quant à la seconde phase, elle amènera à distinguer parmi les variables influençant la qualité celles qui sont pleinement contrôlées (teneur des alliages, par exemple) de celles que subit le producteur (modes d'utilisation du produit par la clientèle) ou qu'il peut difficilement modifier (e.g. décisions d'exportation vers des pays aux conditions climatiques plus ou moins bien adaptées au produit).

Le plan d'expérience, en tant qu'outil statistique, intervient dans la première phase mais il importe de savoir, dès le début, si les variables (potentiellement) explicatives sont essentiellement contrôlées ou subies.

1. Cadre de l'étude

Chercher à “expliquer” une variable quantitative Y (indicateur de qualité, par exemple) à l'aide de n variables x_1, \dots, x_n (reflétant par exemple des conditions de fabrication ou d'emploi) conduit, en général, à rechercher une fonction telle que :

$$Y = f(x_1, \dots, x_n; \theta) + u(x_1, \dots, x_n) \quad (1)$$

où θ est un paramètre (vectoriel) et $u(x_1, \dots, x_n)$ une perturbation aléatoire (reflet en particulier des variables explicatives non prises en compte); f est la fonction de réponse.

Une telle modélisation pose évidemment plusieurs problèmes.

1.1 Problèmes généraux

1.1.1 Plan d'expérience et ajustement

a) Choix de la fonction d'ajustement

La détermination de la relation (1) passe par le choix d'une classe paramétrée \mathcal{F} de fonctions f . Dans les cas les plus favorables, ce choix peut se fonder sur une modélisation conduisant, par exemple, à des équations d'évolution dont la résolution fournira la classe \mathcal{F} . Les paramètres seront alors déterminés par “calage” sur des données.

Exemple :

Souhaitant prévoir l'évolution de $Y(t)$, stock d'équipements téléphoniques des ménages dans un pays à la date t , on peut constater que la demande instantanée, supposée représentée par la dérivée $\frac{dY(t)}{dt}$, dépend du nombre de ménages non encore équipés (c'est le “marché potentiel”) mais aussi du nombre de ménages déjà équipés (qui exercent, pour un équipement de communication, un effet d'entraînement). Si N est le nombre total de ménages, une modélisation du type :

$$\frac{dY(t)}{dt} = \alpha [N - Y(t)]^\beta [Y(t)]^\gamma \quad (\alpha, \beta, \gamma \text{ paramètres réels}) \quad (2)$$

conduira, par intégration, à une classe de fonctions à 4 paramètres à déterminer par ajustement aux observations.

Dans les cas les plus fréquents, on se contentera d'une approche empirique fondée sur des données (d'observation spontanée ou d'expérience) et sur une connaissance antérieure ou intuitive, le choix de la classe \mathcal{F} devant être ultérieurement validé par des tests de qualité de l'ajustement.

b) *Choix du plan d'expérience*

Une fois choisie la classe \mathcal{F} de fonctions, l'ajustement vise donc à estimer le paramètre θ en utilisant les données disponibles.

Mais lorsqu'il s'agit de données expérimentales, comment les recueillir? Plus précisément, comment choisir, si cela est possible, les jeux de valeurs des variables x_1, \dots, x_n pour lesquels on observera les valeurs prises par la variable Y ? Telle est la *question centrale* qu'envisage la théorie des plans d'expérience : une expérience e_i comportera le choix d'un n -uplet (x_{1i}, \dots, x_{ni}) appelé *combinaison expérimentale* et permettra l'obtention d'un résultat expérimental Y_i .

Interprétation géométrique : dans le cas où toutes les variables intervenant sont réelles, on se place dans l'espace euclidien \mathcal{R}^{n+1} rapporté à un repère dont les axes correspondent aux variables x_1, \dots, x_n, Y . A tout n -uplet (x_{1i}, \dots, x_{ni}) correspond un point m_i du sous-espace \mathcal{R}^n engendré par les variables contrôlées (x_1, \dots, x_n) et un point M_i de \mathcal{R}^{n+1} dont m_i est la projection (parallèlement à \vec{oy}) et dont la coordonnée Y_i constitue le résultat expérimental (fig. 1).

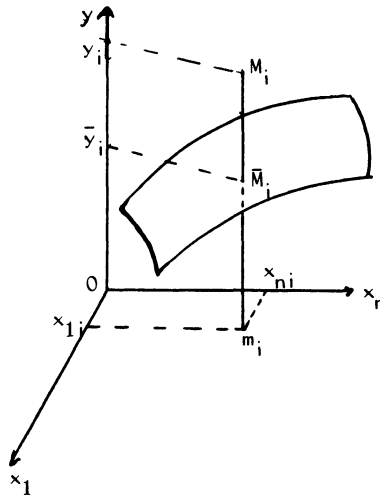


FIGURE 1

Si la perturbation u est d'espérance nulle, plus précisément si :

$$\forall (x_1, \dots, x_n), E[u(x_1, \dots, x_n)] = 0 \quad (3)$$

le point \bar{M} de coordonnées $(x_1, \dots, x_n, \bar{y})$ décrira une "surface de réponse" $s(\theta)$ d'équation :

$$\bar{y} = f(x_1, \dots, x_n; \theta) \quad (4)$$

$f(\cdot; \theta)$ [resp. $s(\theta)$] étant l'élément générique de la classe prédéfinie \mathcal{F} [resp. de la classe φ de surfaces associée à \mathcal{F}] et θ étant à estimer ultérieurement par ajustement.

Choisir un plan d'expérience \mathcal{E} revient à choisir I points m_i (i.e. I combinaisons expérimentales e_i) dont on s'efforcera qu'ils vérifient des propriétés déterminées. Dans cette optique, le plan d'expérience précède l'ajustement mais nous pouvons aussi, sur le conseil de N.R. DRAPER [18,p 77] "fit (say) a linear regression (then) use an experimental design which will protect us against the possibility that the true model is a quadratic or higher polynomial".

1.1.2 Qualités attendues d'un plan d'expérience

Un plan d'expérience doit rendre compte de l'influence des facteurs sur la variable – cible Y . Plus précisément, on distingue (une fois l'“effet moyen” évalué) :

- les *effets principaux* des différents facteurs, c'est-à-dire l'influence X_j de chacune des modalités d'une variable x_j , toutes modalités des autres variables confondues,
- les *effets d'interactions* d'ordre k ($1 \leq k \leq n - 1$) c'est-à-dire les effets associés de $k + 1$ (parmi les n) facteurs, toutes modalités des autres facteurs confondues et déduction faite des effets principaux de ces $k + 1$ facteurs et de leurs interactions d'ordre inférieur.

Ces effets “se lisent” sur la fonction de réponse f , par projection sur le plan $(\vec{0y}, \vec{0x}_j)$ pour l'effet principal X_j du facteur x_j , par projection sur le sous-espace $(\vec{0y}, \vec{0x}_j, \vec{0x}_k)$ si l'on désire connaître l'interaction (d'ordre 1) $X_j X_k$ etc...

Certaines modélisations mettent clairement en évidence les différents effets postulés comme nous en verrons un exemple ultérieurement.

Dès que l'un des facteurs contrôlés est continu, il devient évidemment impossible d'adopter un plan complet, c'est-à-dire de retenir toutes les combinaisons expérimentales. Même dans le cas où tous les facteurs ne comportent qu'un nombre fini de modalités, un choix s'impose souvent :

- on ne peut retenir que des combinaisons expérimentales techniquement et économiquement réalisables, certaines modalités de différents facteurs pouvant être incompatibles : on construit donc un “*plan incomplet*” (ou “*creux*”) et non un “*plan factoriel complet*”,
- parmi ces combinaisons, seront privilégiées celles qui, dans leur ensemble, renseigneront bien sur la fonction de réponse $f(., \theta)$, sur la perturbation u et permettront de déterminer de bonnes prévisions de Y , en particulier pour les configurations $(x_1 \dots x_n)$ “intéressantes”, c'est-à-dire, en général, celles qui conduisent à des valeurs souhaitables de la qualité Y .

Donnons pour l'instant deux exemples des problèmes évoqués ici :

Exemples :

- a) **Régression linéaire simple** : x_1 est la variable contrôlée, Y la variable à expliquer. \mathcal{F} est l'ensemble des fonctions affines du paramètre $\theta = (\beta_0, \beta_1)$:

$$f(x_1; \theta) = \beta_0 + \beta_1 x_1; \quad Y = f(x_1; \theta) + u$$

Sous les hypothèses classiques concernant la perturbation u , il existe un estimateur efficace $\hat{\theta} = (\hat{\beta}_0, \hat{\beta}_1)$ de θ construit à partir de l'échantillon de taille n :

$$(x_{11}, Y_1), \dots, (x_{1i}, Y_i), \dots, (x_{1n}, Y_n)$$

(les x_{1i} étant supposés mesurés exactement et connus).

On sait que l'intervalle de confiance (de niveau $1 - \alpha$) pour la prévision de Y , lorsque $x_1 = x_{10}$, est donné (à l'aide de la table de Student à $n - 2$ ddl) par la formule :

$$\hat{\beta}_0 + \hat{\beta}_1 x_{10} \pm \frac{t(n-2)s}{1-\alpha/2} \sqrt{1 + \frac{1}{n} + \frac{(x_{10} - \bar{x}_1)^2}{\sum_i (x_{1i} - \bar{x}_1)^2}} \quad (5)$$

$$\text{où : } \bar{x}_1 = \frac{1}{n} \sum_{i=1}^n x_{1i}$$

$$\text{et : } s = \left[\sum_i e_i^2 / (n-2) \right]^{1/2} \quad \left(e_i, \text{résidus du modèle : } e_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i}) \right)$$

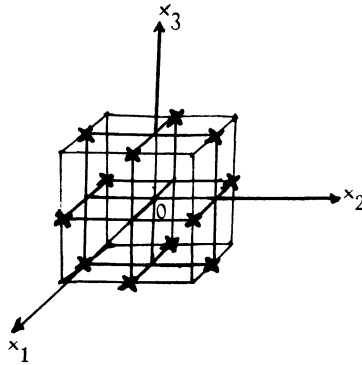
Il apparaît ainsi que la *précision* de la prévision est d'autant plus grande que x_{10} est proche du point moyen \bar{x}_1 . La construction du plan d'expérience ($x_{11} \dots x_{1n}$) est donc guidée par la connaissance des valeurs que l'on pense assigner (à la conception du produit) à la variable de contrôle x_1 (la difficulté étant que cette connaissance précise est justement subordonnée, en général, aux résultats des expériences).

- b) **Plan de type BOX-BEHNKEN** [5]

Supposons qu'interviennent 3 facteurs x_1, x_2, x_3 , chacun prenant l'une des 3 valeurs $-1, 0, +1$. Le plan à 12 combinaisons expérimentales (parmi les 27 possibles), représenté fig.2, est préconisé dans la mesure où il permet d'explorer la courbure de la surface de réponse et les influences couplées (autant de combinaisons expérimentales font intervenir chacune des modalités de chacun des facteurs).

1.1.3. Plan d'expérience et analyse de la variance

Un plan d'expérience doit également se prêter commodément à une analyse de variance : qu'est-ce à dire ? Pour mettre en évidence l'influence des facteurs (individuellement ou en association) sur la variable à expliquer Y , il est naturel d'en étudier la dispersion, - la variance dès lors qu'il s'agit d'une variable réelle.



(Les 12 combinaisons choisies parmi les 27 possibles sont marquées d'une croix x).

FIGURE 2

Conformément à l'étymologie, l'analyse de la variance consiste à décomposer celle-ci en une somme de termes interprétables en tant que variances d'effets principaux ou d'interactions.

Ces effets s'évalueront par des *contrastes*, c'est-à-dire des combinaisons linéaires de résultats expérimentaux y_1, y_2, \dots, y_k de la forme :

$$\sum_{i=1}^k \lambda_i y_i \quad \text{les } \lambda_i, \text{ réels, vérifiant } \sum_{i=1}^k \lambda_i = 0$$

Deux contrastes

$$\sum_i \lambda_i y_i \quad \text{et} \quad \sum_i \mu_i y_i$$

seront dit *orthogonaux* si $\sum_i \lambda_i \mu_i = 0$.

Le recours à des contrastes orthogonaux permet des décompositions de la variance en sommes de carrés (avec disparition des termes "rectangles"). Dans les modèles à effets additifs et sous les hypothèses "classiques" (perturbations u gaussiennes, non corrélées et à variance constante), de telles décompositions facilitent beaucoup la recherche des effets significatifs par application du test de Fisher-Snedecor.

1.1.4 Fractionnement d'un plan

Il arrive que toutes les expériences d'un plan ne puissent être effectuées dans les mêmes conditions : différences de lieux (par exemple essais "multicentriques" dans le domaine biomédical), de climat (par exemple essais in situ dans le domaine du BTP) etc ... Cela amène à fractionner le plan, c'est-à-dire à partitionner, en *blocs*, l'ensemble des expériences qu'il comporte. Cependant l'estimation de certains

effets peut s'en trouver compromise du fait qu'on ne saura plus si les disparités constatées entre les résultats expérimentaux sont dues à certains facteurs ou aux différences entre blocs : c'est le "*confounding*", ou confusion d'effets des facteurs avec ceux dûs aux différences entre blocs.

1.1.5 Plans incomplets

C'est le cas où toutes les combinaisons expérimentales envisageables ne sont pas mises en oeuvre. En fonction des choix effectués, non seulement, comme précédemment, l'estimation de certains effets sera sacrifiée, mais - comme nous le verrons plus précisément dans le cas des plans factoriels réduits - celle des autres effets peut s'en trouver détériorée au point que plusieurs effets recevront la même estimation.

La connaissance des effets que l'on accepte de sacrifier (soit parce qu'on les juge négligeables, soit parce qu'on les connaît déjà) guide dans la construction du plan, mais il ne faut pas perdre de vue qu'un plan construit à partir du sacrifice (consenti) de certaines effets sacrifie parfois automatiquement un certain nombre d'autres effets qu'il convient de répertorier pour savoir si le plan est acceptable.

Laissant de côté, pour le moment, les plans incomplets construits comme "fractions" de plans factoriels complets, nous présenterons brièvement quelques plans incomplets classiques :

- en carrés latins
- en carrés gréco-latins
- en blocs randomisés
- en blocs incomplets équilibrés

a) Carrés latins

Une variable réelle Y est susceptible d'être expliquée par 3 facteurs a , b , c prenant chacun n modalités ($a_i, i = 1 \dots n$; $b_j, j = 1 \dots n$; $c_k, k = 1 \dots n$).

Un plan en carré latin consiste à choisir, parmi les n^3 combinaisons expérimentales, n^2 combinaisons; c'est-à-dire n^2 triplets (a_i, b_j, c_k) tels que, parmi eux, chaque couple (a_i, b_j) , (b_j, c_k) , (a_i, c_k) se retrouve 1 fois exactement. Un tel plan peut être commodément représenté par un carré à n^2 éléments entiers (les indices), chaque entier figurant une fois (et une fois seulement) par ligne et par colonne; ainsi pour $n = 4$, le carré suivant :

1	2	3	4
2	3	4	1
3	4	1	2
4	1	2	3

obtenu simplement par permutations circulaires de la première ligne, traduit, si les numéros de lignes (resp. de colonnes, resp. les éléments du tableau) représentent les indices des modalités du facteur a (resp. b , resp. c), que l'on observera, par exemple,

la valeur y_{234} correspondant à la valeur de la variable Y pour la combinaison expérimentale (a_2, b_3, c_4) , parce que 4 figure à la ligne 2, colonne 3.

Un tel plan incomplet réalise une économie par rapport au plan factoriel complet (n^2 combinaisons au lieu de n^3) mais il ne répond bien qu'à un modèle additif sans interaction.

Notons qu'à n fixé, on a le choix entre $n!(n-1)!$ carrés latins distincts (cf plus loin en 1.2.2 un exemple de critère de choix).

b) Carrés gréco-latins

Quand on passe à 4 facteurs a, b, c, d présentant chacun le même nombre n de modalités, il est avantageux d'envisager encore un plan à n^2 combinaisons expérimentales (a_i, b_j, c_k, d_l) vérifiant les propriétés suivantes :

1. le plan est en carré latin vis-à-vis du facteur c et du facteur d ;
2. si l'on "superpose", les deux carrés, chaque couple (c_k, d_l) apparaît une fois (et une fois seulement) : on dit alors que les deux carrés sont *orthogonaux*, ou que l'ensemble obtenu constitue un carré gréco-latin (les modalités des facteurs c et d étant parfois représentées par des lettres, respectivement latines et grecques).

L'économie procurée par les plans en carrés mutuellement orthogonaux a motivé d'actives recherches jusque vers 1960, en dehors de considérations "ludiques" : on savait, depuis le siècle dernier, qu'il n'existait pas de carré gréco-latin pour $n = 2$ et $n = 6$ (cf Problème dit "des 36 officiers", 6 dans chaque arme et de chaque grade, qu'il est impossible de disposer en carré 6x6, chaque arme et chaque grade étant représentés dans chaque ligne et chaque colonne). EULER crut qu'il n'existait aucun carré gréco-latin pour n entier non multiple de 4. C'est seulement en 1959 que R.C. BOSE et al. établirent qu'il n'en est rien (sauf pour $n = 2$ et $n = 6$) [3].

On peut envisager, si l'on a $K + 2$ facteurs à n modalités chacun, un plan en K carrés latins d'ordre n , mutuellement orthogonaux (il en existe, aux 2 exceptions citées près, et leur nombre tend même vers l'infini avec n [6]), mais la forte économie réalisée se paie d'une perte d'information importante sur les effets des facteurs.

Exemple : $n = 4$ Les carrés suivants sont orthogonaux 2x2 :

1	2	3	4	1	2	3	4	1	2	3	4
3	4	1	2	4	3	2	1	2	1	4	3
2	1	4	3	3	4	1	2	4	3	2	1
4	3	2	1	2	1	4	3	3	4	1	2

c) Plan en blocs randomisés

Exemple : Une société de construction automobile dispose de 5 usines que l'on désire comparer en examinant des véhicules selon 7 critères donnant chacun lieu

à attribution d'une note Y . On prélève 7 véhicules dans la production de chaque usine, soit 35 véhicules que l'on examine chacun selon l'un des 7 critères.

Deux facteurs sont donc envisagés :

- le facteur “critère” A , prenant 7 modalités,
- le facteur “usine” B , prenant 5 modalités.

A chacune des modalités de B correspond un *bloc* à 7 cellules (les véhicules qui proviennent de l'usine associée à cette modalité de B).

Si chaque bloc est constitué d'une permutation des 7 critères tirée au hasard parmi les $7!$ permutations possibles, on a un plan d'expérience en *blocs randomisés*. D'une manière générale, si le facteur A a t modalités, chaque bloc devra comporter t unités expérimentales, les affectations étant aléatoires (tirées au sort parmi les $t!$ possibles en utilisant, par exemple, les tables de COCHRAN et COX [7]) et indépendantes d'un bloc à l'autre.

d) Plan en blocs incomplets

Il arrive que l'on ne souhaite pas avoir par bloc un nombre de cellules égal au nombre de modalités du facteur A (ou même que ce soit impossible : si les cellules d'un même bloc représentent les roues d'une automobile, leur nombre sera limité à 4 alors que les modalités d'un facteur “nature du pneumatique utilisé” pourront être plus nombreuses).

On constitue alors un *plan en blocs incomplets*.

- t est le nombre de *modalités* du facteur (encore appelées “traitement” - en souvenir d'exemples agronomiques)
- b est le nombre de *blocs*
- k est le nombre de *cellules* (ou unités expérimentales) de chaque bloc
- r est le nombre de *répétitions* de chaque modalité dans l'ensemble des expériences.

Il en résulte que si k (resp. r) est constant dans chaque bloc (resp. pour chaque modalité), le nombre total d'observations est égal à

$$rt = bk \quad (6)$$

ce qui constitue donc une relation imposée entre les paramètres du plan d'expérience. L'analyse est techniquement facilitée si chaque paire de modalités apparaît dans *le même nombre λ de blocs*. Un tel plan en blocs incomplets est dit “*équilibré*”.

Exemple : Chaque bloc représente les 4 roues d'un véhicule : $k = 4$. Il y a 7 véhicules : $b = 7$; 7 modalités pour les pneumatiques : $t = 7$. Il en résulte : $r = 4$:chaque modalité sera répétée 4 fois. Voici un plan possible.

n° de bloc	modalités de pneumatique intervenant dans chacune des 4 roues			
1	1	2	3	4
2	5	6	3	4
3	5	7	2	4
4	5	7	1	3
5	7	1	6	4
6	5	1	6	2
7	7	6	2	3

Comme, de surcroît, chaque paire de modalités apparaît dans le même nombre de blocs ($\lambda = 2$), ce plan est *équilibré*.

On établit que $\lambda(t-1) = r(k-1)$ et, plus difficilement, que $b \geq t$ et donc que $r \geq k$.

1.2 Problèmes spécifiques

Des difficultés particulières surviennent dans la construction et l'exploitation des plans d'expérience selon la nature de celles-ci ou les objectifs poursuivis. Nous en mentionnerons ici quelques-unes.

1.2.1 Plan d'expérience en simulation

La simulation est une méthode numérique, expérimentale, d'étude de situations relativement complexes. A ce titre elle requiert la construction de plans d'expérience pouvant comporter de nombreux facteurs, ce qui interdit très vite, malgré la rapidité du calcul électronique, d'essayer toutes les combinaisons de modalités : vingt facteurs à cinq modalités chacun conduiraient déjà à 5^{20} jeux de valeurs d'entrée; même si le traitement de chacun d'eux ne requerrait qu'une microseconde, le temps de calcul nécessaire excèderait mille jours. Procéder à des essais au hasard - même nombreux, en apparence - constitue une exploration parfois exagérément sommaire d'un univers très vaste. Que dire quand interviennent des centaines de facteurs? Des techniques d'agrégation de facteurs et de balayage des groupes ainsi constitués ont été proposées [15].

1.2.2 Enrichissement d'un plan "classique"

Il arrive que se présente l'opportunité d'ajouter quelques expériences à un plan incomplet classique. Cette perspective peut susciter un critère de choix entre plans en apparence équivalents. Reprenons ainsi l'exemple b (1.1.2), celui de 3 facteurs à 3 modalités - 1, 0, 1. On peut construire 12 carrés latins, chacun comportant 9 expériences. Or 4 d'entre eux nécessitent 6 expériences additionnelles pour couvrir le plan de Box-Behnken (mentionné ci-dessus), tandis que les 8 autres en nécessitent 10. Certains carrés latins apparaissent ainsi à cet égard "meilleurs que d'autres" [13].

1.2.3 Plan d'expérience et stabilité de la qualité

Parmi tous les facteurs concourant à la qualité d'un produit, il convient, on l'a dit, de distinguer deux catégories :

- les facteurs de conception et de fabrication dont on peut assez largement choisir les modalités dans une perspective d'optimisation : ce sont des facteurs "internes";
- les facteurs d'utilisation (conditions climatiques, comportement des utilisateurs potentiels ...) dont les modalités sont davantage subies que choisies, (facteurs "externes").

Or ces dernières présentent une forte diversité, en particulier en raison de la "mondialisation" des marchés.

Il convient alors d'évaluer l'influence de ces modalités sur la qualité et d'explorer l'hétéroscédasticité éventuelle des perturbations correspondantes (i.e. la variabilité de la variance de $u(x_1, \dots, x_n)$ dans (1)).

Le problème est le suivant : pour certaines modalités de facteurs internes (contrôlés), la qualité peut être plus ou moins variable selon les modalités des facteurs externes (subis). Préférence pourrait alors être donnée à des configurations dont la qualité ne serait pas optimale en espérance mathématique, mais présenterait l'avantage d'une faible variance, dès lors que serait privilégiée la *stabilité des performances*.

Ce point de vue n'est pas nouveau : il est ainsi émis en 1961 par R.C. BOSE (in [18] p. 285) : "It would be very interesting to consider designs, for which the maximum variance of the estimated response over a given region of interest should be least possible".

2. Etude d'un cas : le plan factoriel 2^n

Nous reprendrons ici, de manière plus précise, les concepts et problèmes évoqués plus haut, à propos d'un cas particulier mais d'application assez fréquente : le plan factoriel 2^n .

Il convient d'évaluer l'influence de n facteurs :

$$a, b, c, \dots k$$

sur une variable réelle Y .

On note :

- l'influence "moyenne" M
- les influences individuelles ou *effets principaux* : $A, B, \dots K$
- les influences combinées ou *interactions* d'ordre 1 : AB, AC, BC, \dots ; d'ordre 2 : $ABC, ABD, \dots BCD, \dots, IJK$; d'ordre supérieur jusqu'à celle d'ordre $n - 1$: $ABC \dots K$.

Nous supposons *ici* que chaque facteur ne prend que 2 modalités : par exemple, dans un alliage, le facteur "chrome" c prendra les modalités "présence" ou "absence". On peut donc concevoir 2^n combinaisons expérimentales.

On note traditionnellement :

(1) : la valeur de Y pour la combinaison correspondant à l'absence de tous les facteurs :

a (resp. b, c, \dots, k) : la valeur de Y pour la combinaison ne comportant la présence que du facteur a (resp. b, c, \dots, k)

ab : la valeur de Y pour la combinaison ne comportant que la présence des facteurs a et b

$abc\dots k$: la valeur de Y pour la combinaison comportant la présence de tous les n facteurs.

Comment évaluer les différents effets à partir de ces valeurs expérimentales ?

2.1 Exemples : plans 2^2 et 2^3

Dans le plan 2^2 , on évalue :

- la moyenne $M = \frac{1}{4} [(1) + a + b + ab]$
- l'effet principal $A = \frac{1}{2} [(a - (1)) + (ab - b)]$
- l'effet principal $B = \frac{1}{2} [(b - (1)) + (ab - a)]$
- l'interaction $AB = \frac{1}{2} [(ab - a) - (b - (1))]$
 $= \frac{1}{2} [(ab - b) - (a - (1))] = BA$

(ainsi l'interaction AB s'interprète comme différence entre les influences de b selon que a est présent ou non).

Notons que les différents effets sont estimés par des *contrastes* orthogonaux 2 à 2.

En assimilant formellement (1) à l'entier 1, on constate les *analogies* :

$$M = \frac{1}{4} (a + 1)(b + 1)$$

$$A = \frac{1}{2} (a - 1)(b + 1)$$

$$B = \frac{1}{2} (a + 1)(b - 1)$$

$$AB = \frac{1}{2} (a - 1)(b - 1)$$

On constate "empiriquement" que le signe $-$ apparaît dans chaque facteur du second membre où apparaît (en minuscule) une lettre figurant (en majuscule) au premier membre.

Cette règle, dégagée empiriquement, se révèle efficace pour l'évaluation des effets dans le plan 2^3 .

Celui-ci comporte 3 facteurs à 2 modalités et donc 2^3 combinaisons donnant à Y les 8 valeurs :

$$(1), a, b, c, ab, ac, bc, abc$$

Les différents effets seront estimés comme suit :

. **Moyenne** : $M = \frac{1}{8} [(1) + a + b + c + ab + ac + bc + abc]$

(formellement) $= \frac{1}{8} (a + 1)(b + 1)(c + 1)$

.Effets principaux

$$A = \frac{1}{4} [(a - (1)) + (ab - b) + (ac - c) + (abc - bc)]$$

(moyenne arithmétique des effets différentiels du facteur a , toutes choses égales par ailleurs);

formellement : $A = \frac{1}{4} (a - 1)(b + 1)(c + 1)$

semblablement : $B = \frac{1}{4} (a + 1)(b - 1)(c + 1)$

$$C = \frac{1}{4} (a + 1)(b + 1)(c - 1)$$

.Interactions d'ordre 1

$$AB = \frac{1}{4} \{[(ab - b) - (a - (1))] + [(abc - bc) - (ac - c)]\}$$

(moyenne arithmétique des influences différentielles du facteur b , sur les effets différentiels du facteur a , toutes choses égales par ailleurs).

.Construction du tableau par colonnes successives :

- Colonne “moyenne” : + à toutes les lignes “combinaisons”
- Colonne “effet principal X ” :
 - + pour les combinaisons expérimentales comportant la lettre x
 - pour les combinaisons ne la comportant pas.
- Colonnes “interaction $XY...$ ” : elles seront constituées de proche en proche à partir des colonnes $X, Y, ...$ par application de la règle classique des signes.

. Construction directe du tableau (c’est-à-dire détermination des signes figurant dans une colonne sans avoir déterminé les autres colonnes) : on applique la règle suivante :

- Règle d’estimation d’un effet (règle de parité)

On considère la parité de l’effet en fonction de son nombre de lettres. Une combinaison expérimentale comportant en commun avec lui des lettres en nombre de *même parité* reçoit le signe + . Sinon, elle reçoit le signe - . Conventionnellement, 0 est considéré comme un nombre pair.

Il est facile de voir que son application permet de retrouver les résultats précédents (ce qui n’en constitue pas une démonstration).

Interprétation géométrique

Dans l’espace euclidien de dimension 2^3 , on considère le vecteur des observations expérimentales :

$$\vec{y} = ((1) a b c ab ac bc abc)$$

dont la représentation matricielle en colonne sera notée y (sa transposée, en ligne, étant notée y').

Si l’on note le vecteur des effets :

$$\vec{E} = (8M 4A 4B 4C 4AB 4AC 4BC 4ABC)$$

et ses représentations matricielles E en colonne et E' en ligne, on voit que l’on passe de \vec{y} à \vec{E} par une transformation linéaire de matrice J' transposée d’une matrice J d’ordre 2^3 , obtenue à partir de la matrice de signes précédente, en remplaçant les + par 1 et les - par - 1 :

$$E' = y'J \text{ soit } E = J'y \quad (7)$$

Il importe de remarquer que la matrice J ainsi définie a ses colonnes mutuellement orthogonales et que plus précisément la matrice :

$$\tilde{J} = \frac{1}{2^{3/2}} J \quad (8)$$

est orthogonale : $\tilde{J}^{-1} = \tilde{J}'$

Il en résulte :

$$\begin{aligned} y &= (J^{-1})' E = \left[(2^{3/2} \tilde{J})^{-1} \right]' E = \left[\frac{1}{2^{3/2}} \tilde{J}^{-1} \right]' E \\ &= \left[\frac{1}{2^{3/2}} \tilde{J}' \right]' E = \frac{1}{2^{3/2}} \tilde{J} E \\ y &= \frac{1}{2^3} J E \end{aligned} \quad (9)$$

On reconstruit donc facilement les valeurs expérimentales à partir des effets. Nous admettons : les résultats précédents s'étendent au cas d'un entier quelconque n :

$$y = \frac{1}{2^n} J E \quad \text{où } \vec{E} = (2^n M, 2^{n-1} A, 2^{n-1} B, \dots, 2^{n-1} ABC \dots K) .$$

2.3 Modèle probabiliste sous-jacent

Posons :

$$Y = Z\beta + u \quad (10)$$

Avec :

$$Z = \frac{1}{2^n} J$$

u : perturbation aléatoire vérifiant les hypothèses classiques ($u \in \mathcal{N}(0; \sigma^2 I_{2n})$).

β : vecteur dont les 2^n composantes désignent des paramètres certains mais inconnus.

A partir de la réalisation du vecteur aléatoire Y , on peut construire le BLUE (estimateur des moindres carrés) du vecteur β :

$$\begin{aligned} \hat{\beta} &= (Z'Z)^{-1} Z'Y \\ &= \left(\frac{1}{2^{2n}} J'J \right)^{-1} \cdot \frac{1}{2^n} J'Y \end{aligned} \quad (11)$$

utilisant (8) :

$$\hat{\beta} = \left(\frac{2^n}{2^{2n}} \tilde{J}'\tilde{J} \right)^{-1} \cdot \frac{1}{2^n} J'Y$$

soit en raison de l'orthogonalité de \tilde{J} :

$$\hat{\beta} = J'Y \quad (12)$$

Le vecteur \vec{E} des effets estimés (tel qu'il a été ci-dessus défini) apparaît, puisque $E = J'y$, comme l'estimation efficace du vecteur de paramètres β du modèle probabiliste ci-dessus.

Analyse de la variance

Le vecteur \vec{E} des effets a donc comme composantes $2^{n-1}X$, X étant un effet principal ou une interaction.

Si chaque combinaison expérimentale est répétée r fois, X sera estimé par un contraste C_x différence entre :

$$\begin{aligned} & r \cdot 2^{n-1} \text{ résultats affectés du signe } + \\ & r \cdot 2^{n-1} \text{ résultats affectés du signe } - \end{aligned}$$

divisé par $r \cdot 2^{n-1}$:

$$\hat{X} = \frac{C_x}{r \cdot 2^{n-1}} \quad (13)$$

Si σ^2 est la variance de la perturbation u affectant les observations :

$$\text{Var} \hat{X} = \frac{\text{Var} C_x}{r^2 \cdot 2^{2(n-1)}} = \frac{2r \cdot 2^{n-1} \sigma^2}{r^2 \cdot 2^{2(n-1)}} = \frac{\sigma^2}{r \cdot 2^{n-2}} \quad (14)$$

Si s^2 estime sans biais σ^2 , l'effet X sera significatif au niveau α si

$$\begin{aligned} |\hat{X}| > t_{1-\alpha/2}(\mu) \frac{s}{\sqrt{r \cdot 2^{n-2}}} \quad (T(\mu) : \text{v.a. de Student à } \mu \text{ ddl} ; \quad (15) \\ P(|T(\mu)| > t_{1-\alpha/2}(\mu)) = \alpha \end{aligned}$$

$$s^2 \text{ peut être calculé comme quotient } \frac{Q_{\text{res.}}}{(r-1)(2^n-1)} \quad (16)$$

$$\text{où} \quad Q_{\text{res.}} = Q_T - \sum_x Q_x - Q_r$$

Q_T : somme des carrés des écarts entre les résultats ($r2^n - 1$ ddl)

$Q_x = \frac{C_x^2}{r \cdot 2^n}$: carré associé à l'effet X (1 ddl)

Q_r : somme des carrés entre répétitions ($r - 1$ ddl)

$$\mu = (r2^n - 1) - (2^n - 1) - (r - 1) = (r - 1)(2^n - 1)$$

2.4. Fractionnement d'un plan 2^n en blocs

Exemple :

Supposons qu'un plan 2^3 soit fractionné en 2 blocs (Z_1, Z_2) (resp Z'_1, Z'_2) de 4 combinaisons chacun.

TABLEAU 2

	8M	4A	4B	4C	4AB	4AC	4BC	4ABC	Z_1	Z_2	Z'_1	Z'_2
(1)	+	-	-	-	+	+	+	-		(1)	(1)	
a	+	+	-	-	-	-	+	+	a		a	
b	+	-	+	-	-	+	-	+	b			b
c	+	-	-	+	+	-	-	+	c			c
ab	+	+	+	-	+	-	-	-		ab		ab
ac	+	+	-	+	-	+	-	-		ac		ac
bc	+	-	+	+	-	-	+	-		bc	bc	
abc	+	+	+	+	+	+	+	+	abc		abc	

S'il y a un effet "bloc" (supposé additif) :

- il sera estimé, dans le cas du fractionnement Z_1, Z_2 , par $\frac{1}{4} C_Z$ où C_Z est le contraste défini par :

$$C_Z = (a + b + c + abc) - ((1) + ab + ac + bc)$$

qui est précisément celui intervenant dans l'estimation de l'effet ABC : cet effet ABC sera dit *confondu* avec l'effet bloc.

- dans le cas du fractionnement Z'_1, Z'_2 interviendra :

$$C_{Z'} = ((1) + a + bc + abc) - (b + c + ab + ac)$$

qui sert également à estimer l'effet BC (qui est donc ici l'effet confondu).

Le fractionnement en 2 blocs consiste à répartir les 2^n combinaisons expérimentales en 2 blocs de 2^{n-1} combinaisons. L'effet bloc ajoute un paramètre au modèle : reprenant l'interprétation géométrique, nous avons donc à décomposer le vecteur \vec{y} (à 2^n composantes) sur $2^n + 1$ vecteurs, ce qui conduit nécessairement à une indétermination puisque leur rang est exactement 2^n .

L'estimation d'un effet se trouve donc sacrifiée : on choisira si possible un effet connu par ailleurs ou présumé négligeable. L'exemple montre comment opérer le fractionnement en fonction de l'effet que l'on a sacrifié : dans un bloc, toutes les combinaisons affectées du signe + dans la colonne de cet effet, dans l'autre celles affectées du signe - .

Plus généralement, le fractionnement en 2^k blocs présente une difficulté supplémentaire (pour $k > 1$) : poursuivant l'exemple ci-dessus, on partitionne Z_1 en Z_{11}, Z_{12} et Z_2 en Z_{21}, Z_{22} comme suit :

TABLEAU 3

	Z_{11}	Z_{12}	Z_{21}	Z_{22}
(1)			(1)	
a	a			
b		b		
c		c		
ab				ab
ac				ac
bc			bc	
abc	abc			

On peut considérer (en examinant avec soin les signes de la matrice du Tableau 2) que les 2 fractionnements successifs ont été engendrés par les sacrifices de l'effet ABC puis de l'effet BC . Cependant, si les effets blocs sont supposés additifs, ce sont tout de même 3 paramètres indépendants qui ont été ajoutés au modèle, ce qui entraîne 3 indéterminations et donc 3 effets confondus : on peut remarquer que l'effet A est maintenant également confondu puisqu'il est estimé, d'après la règle de parité, par :

$$\frac{1}{4} [(a + abc) - (b + c) - ((1) + bc) + (ab + ac)]$$

nos groupements mettant ici en évidence que les effets blocs ne sont pas éliminés.

Par contre, les autres effets sont "estimables", par exemple :

- l'effet B par

$$\frac{1}{4} \left[\begin{array}{cccc} (abc - a) & + & (b - c) & + & (bc - (1)) & + & (ab - ac) \\ (Z_{11}) & & (Z_{12}) & & (Z_{21}) & & (Z_{22}) \end{array} \right]$$

où les effets blocs (supposés additifs) s'éliminent.

- l'effet C par

$$\frac{1}{4} [(abc - a) + (c - b) + (bc - (1)) + (ac - ab)]$$

Comment déterminer l'ensemble des effets confondus par fractionnement en 2^k blocs ?

On constate sur notre exemple ($n = 3$, $t = k = 2$) que 2 des effets confondus engendrent le troisième par multiplication "formelle" :

$$ABC \cdot BC = A \cdot B \cdot B^2 \cdot C^2 = A \text{ en posant } \forall X, X^2 = I \text{ (élément unité) ;}$$

$$\text{et de même : } ABC \cdot A = A^2 BC = I \cdot BC = BC$$

Ce résultat s'étend en fait à k , entier quelconque $< n$:

. k effets sont des *générateurs distincts* si aucun de leurs produits (selon la loi précédente 2 à 2, 3 à 3, ...) ne figure parmi eux.

. Un ensemble de k générateurs distincts permet de réaliser un fractionnement du plan 2^n en 2^k blocs (chacun comportant donc 2^{n-k} combinaisons expérimentales) et engendre par multiplication formelle un groupe de $2^k - 1$ effets qui seront confondus avec les effets blocs dans ce fractionnement.

2.5 Plans réduits 2^{n-k}

Des contraintes diverses (temps, budget) conduisent souvent à renoncer au plan complet 2^n et à ne conserver que certains des blocs définis précédemment par fractionnement.

Exemple : $n = 3$; facteurs a, b, c

Le fractionnement en deux blocs :

$$Z_1 : a, b, c, abc$$

$$Z_2 : (1), ab, ac, bc$$

avait conduit à sacrifier l'estimation de l'effet ABC .

Si l'on se limite au bloc Z_1 , on disposera pour estimer A de 4 résultats seulement qui seront utilisés "au mieux" en calculant :

$$\frac{1}{2}(abc + a - b - c)$$

Or c'est le même contraste qui conduira à estimer l'effet BC .

Les effets A et BC recevront ainsi la même estimation, ce que l'on notera :

$$[A = BC]$$

A et BC seront dits "alias" relatifs à Z_1 .

De même : $[B = AC]$ (car on adoptera : $B = AC = \frac{1}{2}(abc + b - a - c)$)
 $[C = AB]$ ($C = AB = \frac{1}{2}(abc + c - a - b)$)

Si l'on se limite au bloc Z_2 , on trouve :

$$\left. \begin{array}{l} [A = -BC] \quad (A = \frac{1}{2}(ab + ac - bc - (1)) = -BC) \\ [B = -AC] \quad (B = \frac{1}{2}(ab + bc - ac - (1)) = -AC) \\ [C = -AB] \quad (C = \frac{1}{2}(ac + bc - ab - (1)) = -AB) \end{array} \right\} \text{ alias relatifs à } Z_2$$

Les alias d'un effet sont les effets qui s'expriment au moyen du même contraste que l'effet considéré.

Notons que les alias de chaque effet s'obtiennent par multiplication "formelle" de celui-ci par l'effet confondu par le fractionnement ayant conduit au plan réduit considéré :

$$\begin{array}{l} A. ABC = A^2 BC = BC \quad (\text{car } A^2 = I) \\ B. ABC = AC \\ C. ABC = AB \end{array}$$

Plus précisément, le bloc Z_1 correspond à ABC , le bloc Z_2 à $-ABC$.

Les résultats constatés sur cet exemple s'étendent au cas général d'un plan 2^n réduit à 2^{n-k} (k entier, $1 \leq k < n$). Un tel plan est obtenu par fractionnement du plan complet 2^n en 2^k blocs (comportant chacun 2^{n-k} combinaisons expérimentales).

Ce fractionnement correspond à un groupe générateur de $2^k - 1$ effets, lui-même engendré à partir de k effets générateurs distincts.

Parmi les $2^n - 1$ effets (principaux ou d'interaction), on en comptera $2^k - 1$ qui seront "confondus" avec l'"effet bloc". De plus, chaque effet X non confondu recevra la même estimation que chacun des effets obtenus par multiplication "formelle" de X par chacun des $2^k - 1$ effets confondus.

Nous avons ainsi $2^n - 2^k$ effets non confondus répartis en classes comportant chacune 2^k alias (effets recevant simultanément la même estimation). Il y a donc $2^{n-k} - 1$ classes.

Les contrastes C_X s'obtiendront par la règle de parité, les estimations étant obtenues en les divisant par 2^{n-k-1} .

Exemple :

$$n = 5 : 5 \text{ facteurs } a, b, c, d, e,$$

$$k = 2 : \text{réduction à un bloc sur quatre,}$$

$$2^5 = 32 \text{ combinaisons expérimentales dans le plan complet.}$$

Si l'on accepte de sacrifier la connaissance de deux interactions du second ordre ABC et CDE , l'interaction du troisième ordre :

$$ABC \cdot CDE = AB C^2 DE = ABDE$$

se trouvera automatiquement sacrifiée également.

Comme ABC et CDE sont impairs, l'un des blocs constitué comportera les combinaisons ayant en commun avec ces effets un nombre impair de lettres, soit :

$$c, ad, ae, bd, be, abc, cde, abcde$$

Les effets à estimations simultanées de chaque effet X non confondu ("alias") s'obtiennent par multiplication "formelle" de X par les éléments :

$$ABC, CDE, ABDE$$

du groupe générateur du fractionnement.

Soit :

$$\begin{aligned} [A &= BC = ACDE = BDE] \\ &= \frac{1}{4}(abcde + abc + ad + ae - cde - bd - be - c) \\ [B &= AC = BCDE = ADE] \\ [C &= AB = DE = ABCDE] \\ [D &= ABCD = CE = ABE] \\ [E &= ABCE = CD = ABD] \\ [AD &= BCD = ACE = BE] \\ [AE &= BCE = ACD = BD] \end{aligned} \quad (17)$$

$$(2^{n-k} - 1 = 7 \text{ classes comportant chacune } 2^k = 4 \text{ alias}).$$

Remarques :

- 1) On évitera en général, dans la construction d'un plan réduit par fractionnement, les générateurs distincts égaux à des effets principaux (ou même à des interactions d'ordre 1) ou *qui en engendrent par multiplication formelle* (du type ABC et BC qui engendrent évidemment A). De ce fait, les effets principaux ne recevront d'estimations simultanées qu'avec des interactions d'ordre ≥ 1 : de tels plans sont dits de résolution III (par exemple plan 2^5 fractionné par $ABCDE$ et EFG et donc ne confondant que les effets : $ABCDE, EFG, ABCDFG$).

Un effet principal n'aura ainsi comme alias qu'une interaction d'ordre 1 au moins. Notons que le même plan 2^5 conduira par le groupe générateur :

$$\{ABCD, DEFG, ABCD.DEFG = ABCEFG\}$$

à des plans de résolution IV ou chaque effet principal n'aura comme alias que des interactions d'ordre ≥ 2).

- 2) On peut être amené à répéter des combinaisons expérimentales alors même que l'on a dû se contenter de plans réduits : les répétitions permettent d'évaluer la variance de l'"erreur de mesure". De plus, si elles sont convenablement réparties en cas de fractionnements multiples, elles permettent,

si certains effets sont confondus dans l'un de ceux-ci, qu'ils ne le soient pas dans les autres, ce qui en rend possibles leurs estimations.

2.6 Généralisation et problèmes connexes

Cette présentation du plan 2^n laisse fatalement l'impression que son traitement bénéficie d'heureuses surprises (règle de parité, lien entre générateur de fractionnement et estimations simultanées ...).

D'autres liens renforceront cette impression.

a) Plan factoriel et plan en blocs incomplets équilibrés

A partir d'un plan factoriel 2^n , on peut, pour $s < n$, construire par fractionnement 2^{n-s} blocs (comportant chacun 2^s combinaisons) ce qui créera un groupe générateur de $2^{n-s} - 1$ effets confondus. On distingue, en variant le fractionnement, différents ensembles d'effets confondus en nombre :

$$N_{n,s} = \frac{(2^n - 1)(2^n - 2) \dots (2^n - 2^{n-s-1})}{(2^{n-s} - 1)(2^{n-s} - 2) \dots (2^{n-s} - 2^{n-s-1})} \quad (18)$$

Or, il est possible, à partir de là, de construire un plan en blocs incomplets équilibrés où :

$$t = 2^n - 1 \text{ modalités ("variétés")}$$

$$k = 2^{n-s} - 1 \text{ cellules par blocs}$$

$$b = N_{n,s} \text{ blocs}$$

et donc :

$$r = \frac{bk}{t} = \frac{(2^{n-s} - 1) N_{n,s}}{2^n - 1} = \frac{(2^n - 2) \dots (2^n - 2^{n-s-1})}{(2^{n-s} - 2) \dots (2^{n-s} - 2^{n-s-1})}$$

$$\lambda = r \cdot \frac{k - 1}{t - 1} = \frac{(2^n - 2^2) \dots (2^n - 2^{n-s-1})}{(2^{n-s} - 2^2) \dots (2^{n-s} - 2^{n-s-1})}$$

Exemple : $n = 3$; $s = 1$; 3 facteurs a, b, c ; fractionnement en $2^{n-1} = 4$

$$t = 7 \text{ modalités } x_1 \dots x_7$$

$$k = 3 \text{ cellules par blocs}$$

$$b = N_{3,1} = \frac{(2^3 - 1)(2^3 - 2)}{(2^2 - 1)(2^2 - 2)} = 7 \text{ blocs}$$

$$r = 3 \text{ répétitions}$$

$$\lambda = 1 \text{ (chaque couple de modalités n'apparaît qu'une fois dans un même bloc)}$$

TABLEAU 4

Effets confondus blocs	A	B	C	AB	BC	CA	ABC
1	x_1	x_2		x_4			
2	x_1		x_3			x_6	
3	x_1				x_5		x_7
4		x_2	x_3		x_5		
5		x_2				x_6	x_7
6			x_3	x_4			x_7
7				x_4	x_5	x_6	

(les "creux" des lignes constituent d'ailleurs un autre plan en blocs incomplets équilibrés avec $b = 7, t = 7, k = 4, r = 4, \lambda = 2$).

b) Codes correcteurs

Supposons qu'un canal transmette des 0 et des 1 (ou plus généralement p symboles, p étant un nombre premier), mais avec possibilité d'erreur (substitution d'un symbole à un autre). Comment constituer son "dictionnaire" de façon que le récepteur puisse découvrir le message émis par la source alors même que les "mots" peuvent comporter, à l'arrivée, des lettres erronées - en nombre toutefois limité - ? R. FISHER traite ce problème de "codes correcteurs" en 1942, dans le cas d'une seule erreur (dans un mémoire sur le fractionnement [11]), avant HAMMING (1950). En 1960, R.C. BOSE réussit à mettre en bijection codes correcteurs et groupes générateurs de fractionnement [2].

Pour généraliser les résultats présentés et se dégager de ce sentiment de coïncidences fortuites, il convient d'utiliser quelques concepts mathématiques d'apparence abstraite; nous en donnons, pour terminer, un très bref aperçu :

- 1) *Corps de Galois* : un exemple simple - et fondamental - de corps de Galois est l'ensemble des classes résiduelles modulo p , p étant un nombre premier fixé. Appelé $GF(p)$ - de l'anglais Galois field - il comporte p éléments que l'on peut noter :

$$0, 1, 2, \dots, p - 1$$

et il est muni de deux lois de composition internes associatives, commutatives, la seconde distributive par rapport à la première :

- une "addition" admettant un élément neutre 0 et chaque élément ayant un opposé (et un seul) : l'addition modulo p .
- une "multiplication" admettant un élément unité, chaque élément (autre que le neutre) ayant un inverse (et un seul) - grâce au caractère premier de p : la multiplication modulo p .

$GF(p)$ est ainsi un corps fini (donc commutatif). Dans la théorie des plans d'expérience, il interviendra quand les facteurs auront chacun p (premier) modalités (mais on peut généraliser au cas de p^m modalités, p premier et m entier quelconque, en définissant un corps de Galois $GF(p^m)$). Le plan factoriel p^n aura ainsi p^n combinaisons expérimentales notées chacune par un "point" :

$$(x_1, x_2, \dots, x_i, \dots, x_n)$$

où x_i , élément de $GF(p)$ désignera la modalité du $i^{\text{ème}}$ facteur.

Dans le cas du plan 2^2 étudié plus haut, l'effet principal A s'obtenait par comparaison des points $(0,0)$, $(0,1)$ avec les points $(1,0)$, $(1,1)$, donc si l'on avait voulu - à ce niveau - compliquer les choses, par comparaison des 2 solutions de l'équation :

$$x_1 = 0$$

avec celles de l'équation :

$$x_1 = 1$$

Pour l'effet B , comparaison des solutions des équations :

$$x_2 = 0 \quad \text{et} \quad x_2 = 1$$

Pour l'interaction AB , comparaison des points $(0,0)$, $(1,1)$ avec les points $(1,0)$, $(0,1)$, - donc comparaison des solutions de l'équation :

$$x_1 + x_2 = 0 \quad (\text{mod. } 2)$$

avec celles de l'équation :

$$x_1 + x_2 = 1 \quad (\text{mod. } 2)$$

C'est la voie de la généralisation : chaque équation de la forme :

$$\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n = k \quad (\text{mod. } p) \quad (19)$$

où coefficients non tous nuls $(\alpha_1, \dots, \alpha_n, k)$ et inconnues (x_1, \dots, x_n) sont dans $GF(p)$ admet, dans ce corps, p^{n-1} solutions. En les comparant, lorsque k varie de 0 à $p-1$, on obtiendra une évaluation de l'effet $X_1^{\alpha_1} X_2^{\alpha_2} \dots X_n^{\alpha_n}$, qui désigne *une part de l'interaction* entre tous les facteurs d'exposants α_i non nuls : cette notation commode permet de décomposer chaque interaction en effets indépendants entre eux (sous les hypothèses gaussiennes classiques), de degrés de liberté $p-1$ et qui peuvent donner lieu à estimation et test statistiques. Pour s'assurer l'unicité de cette décomposition, on convient de fixer à 1 le premier exposant α_i non nul intervenant dans chaque effet $X_1^{\alpha_1} X_2^{\alpha_2} \dots X_n^{\alpha_n}$. Ainsi dans le cas de 3 facteurs a, b, c , à 5 modalités chacun, les interactions d'ordre 1 seront repérées par les effets suivants :

- . entre a et b : AB, AB^2, AB^3, AB^4 (il serait équivalent d'écrire : $A^\alpha B^\alpha$,
 $A^\alpha B^{2\alpha}, A^\alpha B^{3\alpha}, A^\alpha B^{4\alpha}$ avec $\alpha \in GF(5)$)
- . entre a et c : AC, AC^2, AC^3, AC^4
- . entre b et c : BC, BC^2, BC^3, BC^4

(Ajoutons un détail technique : dans les équations (19), on peut effectivement faire en sorte que le premier α_i non nul, au premier membre, soit égal à 1 : il suffit pour cela de multiplier les 2 membres par un entier convenable, qui existera toujours en raison de la propriété de corps de $GF(p)$).

Si deux effets :

$$A = X_1^{\alpha_1} X_2^{\alpha_2} \dots X_n^{\alpha_n} \quad \text{et} \quad B = X_1^{\beta_1} X_2^{\beta_2} \dots X_n^{\beta_n}$$

sont confondus, les équations :

$$\sum_i \alpha_i x_i = k \pmod{p} \quad \text{et} \quad \sum_i \beta_i x_i = k' \pmod{p} \quad (k, k' \in \{0, 1, \dots, p-1\})$$

ont leurs solutions en commun (dans $GF(p)$). Il en résulte que ces solutions sont également celles des équations :

$$\sum_i (\alpha_i + \lambda \beta_i) x_i = k + \lambda k' \pmod{p}, \quad \forall \lambda \in GF(p)$$

Par suite, l'effet $A B^\lambda$ sera également confondu.

On comprend ainsi pourquoi une multiplication "formelle" intervenait dans le groupe générateur du fractionnement pour $p = 2$ (en particulier, on avait $A^2 = B^2 = \dots = I$ simplement parce que dans $GF(2)$, $2 = 0$).

Exemple : reprenons le cas (mentionné en 2.4) d'un plan 2^3 fractionné en deux blocs Z_1, Z_2 . Le sacrifice de l'estimation de l'effet ABC conduit à rechercher les solutions (entières mod.2, donc 0 ou 1) des équations :

$$x_1 + x_2 + x_3 = 0 \pmod{2} \quad (20)$$

$$x_1 + x_2 + x_3 = 1 \pmod{2} \quad (21)$$

L'équation (20) admet comme solutions les points :

$$(0, 0, 0) = (1) ; (0, 1, 1) = bc ; (1, 0, 1) = ac ; (1, 1, 0) = ab$$

ce sont précisément les combinaisons expérimentales constituant le bloc que nous avons appelé Z_2 .

L'équation (21) admet comme solutions les points associés au bloc Z_1 . Poussons le fractionnement en acceptant de plus le sacrifice de l'effet BC : cela conduit aux équations :

$$x_2 + x_3 = 0 \pmod{2} \quad (22)$$

$$x_2 + x_3 = 1 \pmod{2} \quad (23)$$

et aux différents blocs à constituer, par croisement 2 à 2 des équations; ainsi, le système d'équations $\{(20), (22)\}$:

$$\begin{cases} x_1 + x_2 + x_3 = 0 & \pmod{2} \\ x_2 + x_3 = 0 & \pmod{2} \end{cases}$$

admet comme solutions les points : $(0,0,0)=(1)$ et $(0,1,1)=bc$ ce qui constitue le bloc que nous avons noté Z_{21} . Il est aisé de voir que le système d'équations $\{(20), (23)\}$ conduit au bloc Z_{22} ; $\{(21), (22)\}$ conduit au bloc Z_{11} ; $\{(21), (23)\}$ conduit au bloc Z_{12} .

Notons que la réunion des 2 familles d'équations :

$$x_1 + x_2 + x_3 = k \pmod{2} \quad \text{et} \quad x_2 + x_3 = k' \pmod{2}$$

contient (par soustraction membre à membre) la famille : $x_1 = k'' \pmod{2}$. C'est la famille associée à l'effet A . Le sacrifice de ABC et de BC entraîne celui de A .

- 2) *Géométries finies* : une géométrie euclidienne finie de dimension k consiste en "points" $(x_1, \dots, x_i, \dots, x_k)$, chaque x_i étant un élément d'un corps de Galois. Le lien que nous avons mis en évidence entre l'effet $X_1^{\alpha_1} \dots X_k^{\alpha_k}$ et le point $(\alpha_1, \dots, \alpha_k)$ permet d'interpréter, en termes géométriques, les notions introduites dans l'étude des plans d'expérience, - ainsi que dans celles des codes (par exemple, la détermination d'un "bon" fractionnement ou celle d'un "bon" code correcteur reviendra à construire le maximum de points distincts dans un espace fini, tels que d (entier fixé) quelconques d'entre eux soient linéairement indépendants [2,12]).

Précisons enfin que l'exigence d'unicité de la décomposition des effets qui nous a conduits à remplacer tout vecteur $(\alpha_1 \dots \alpha_n)$ non nul par le vecteur proportionnel $(\lambda \alpha_1 \dots \lambda \alpha_n)$, de manière à donner la valeur 1 à sa première composante non nulle, conduit, de manière équivalente, à préférer à la géométrie euclidienne la *géométrie projective* qui identifie des vecteurs quand ils sont proportionnels entre eux.

Conclusion

La théorie des plans d'expérience se situe ainsi à la rencontre de mathématiques pures (algèbre, géométrie) et appliquées (calcul des probabilités, théorie de l'information, statistique). Née dans les premières années du développement de la statistique moderne, l'intérêt continu qu'elle suscite témoigne de l'importance des problèmes, divers et renouvelés, qu'elle se propose de traiter.

Références

- [1] BOSE R.C. (1950) - Mathematics of factorial designs, *Proceedings of the International Congress of Mathematicians, Vol. I, pp. 543-548.*
- [2] BOSE R.C. (1961) - On some connections between the design of experiments and information theory, *Bull. de l'Inst. Int. de Stat., 38.*
- [3] BOSE R.C., SHRIKHANDE S.S., PARKER E.T. (1960.) - Further results on the construction of mutually orthogonal Latin Squares and the falsity of Euler's conjecture, *Canad. Journ. of Math. 12, pp 189-203.*
- [4] BOX G.E.P., DRAPER N.R. (1959) - A basis for the selection of a response surface

design, *J.A.S.A.*, 54, pp 622-654.

- [5] BOX G.E.P, HUNTER J.S. (1957) - Multifactor experimental designs for exploring response surface, *A.M.S.* 28, pp 195-241.
- [6] CHOWLA S., ERDÖS P., STRAUS E.G. (1960) - "On the maximal number of pairwise orthogonal Latin Squares of a given order", *Canad. Journ. of Math.* 41, pp 204-208.
- [7] COCHRAN W.G., COX G.M. (1950) - Experimental Designs, *J. WILEY*.
- [8] DUGUE D. - Traité de Statistique théorique et appliquée, *tome 2, MASSON*.
- [9] DUGUE D., GIRAULT M. - Analyse de la variance et plan d'expérience, *DUNOD*.
- [10] FISHER R.A. (1935) - The Design of Experiments, *Hafner Publishing Compagny Inc, New-York*, 5^{ème} Ed.,(1949).
- [11] FISHER R.A. (1942) - The theory of confounding in factorial experiments in relation to the theory of groups, *Ann. Eugen. London* 11, pp. 341-353.
- [12] FOATA D. - Sur la construction des plans factoriels fractionnés et certains codes correcteurs, in [18] pp. 137-146.
- [13] HUNTER J.S. (1987) - Are some Latin Squares better than others? in Design, Data and Analysis, pp. 163-169 *C.L. Mallows Ed., J. Wiley*.
- [14] KEMPTHORNE O. (1952) - The Design and Analysis of experiments, *J. Wiley, N.Y.*
- [15] KLEIJNEN J.P.C. (1986) Statistical Tools for Simulation practitioners, *Marcel Dekker, Inc. N.Y.*
- [16] MANN H.B. - Main effects and interactions in factorial designs, in [18] pp. 33-37.
- [17] VESSEREAU A. - Méthodes Statistiques en biologie et en agronomie, *Baillières, Paris*.
- [18] Le Plan d'expériences (1963), in *Colloques Internationaux du CNRS, n° 110, Ed. du CNRS*.