

PLANS D'EXPERIENCE

Préface

A la demande des professionnels de la chimie, la partie statistique du programme de mathématiques du B.T.S. chimie s'enrichit d'un item nouveau : l'initiation aux plans d'expérience. En effet cette technique mathématique est de plus en plus répandue dans l'industrie chimique. Lors de leur stage, certains élèves des sections de techniciens supérieurs se demandent si on leur fait faire des mathématiques ou de la chimie. Une grande partie du temps dans l'entreprise est consacrée à dépouiller des résultats d'expériences et à les combiner pour estimer des paramètres. Ces expériences sont faites selon un plan précis.

Au niveau industriel, la confection et l'utilisation d'un plan d'expérience conforme à un cahier des charges précis, à une modélisation donnée des phénomènes étudiés, requiert des outils mathématiques relativement sophistiqués : algèbre linéaire, algèbre finie, géométrie sur des corps finis, modèle linéaire généralisé de la statistique. Il est évidemment hors de question d'aborder ces questions dans les classes de techniciens supérieurs. En revanche, conformément à l'esprit de l'enseignement des mathématiques dans ces classes, il s'agit, par des exemples simples, de faire comprendre aux élèves quelle est la démarche mise en œuvre. Au delà, la formation continue pourra s'appuyer sur cette formation pour introduire les outils complexes concrètement utilisés par les professionnels et qui sont fournis par des logiciels présents sur le marché.

Les plans d'expérience étant très peu enseignés à l'université, chacun sait que les professeurs de mathématiques ignorent très souvent cette technique. Le présent exposé est destiné à leur donner une première approche. Ce n'est pas un cours à destination des élèves car il déborde évidemment le programme. Une transposition didactique est à effectuer par le professeur ; elle varie suivant le public d'élèves qu'il a devant lui. Acquérir un savoir d'une façon autodidacte est un exercice profitable pour tout enseignant. Le texte qui suit permettra d'amorcer la démarche qui devra se prolonger par la lecture des ouvrages de vulgarisation.

J.-L. PIEDNOIR
I.G.E.N.

Les plans d'expérience

I. De quoi s'agit-il ?

La notion de *plan d'expérience* est ancienne mais l'utilisation systématique de protocoles d'expérience définis à l'avance dans l'industrie et particulièrement dans l'industrie chimique est une chose récente. Pour illustrer la démarche, prenons un exemple. Un industriel gazéifie du bois. Il désire bien entendu avoir le meilleur rendement possible. Pour produire le gaz, il a le choix entre deux types de gazogène, entre deux essences d'arbre : le hêtre ou l'épicéa. Le bois étant coupé, il se demande quelle est la dimension optimale des plaquettes de bois utilisées, quel degré d'humidité des dites plaquettes prendre. Voilà quatre facteurs énoncés ; on pourrait en faire intervenir d'autres : conditions météorologiques par exemple et, pourquoi pas, l'âge moyen de la main d'œuvre.

Face à ce problème, tout individu rationnel commencera à observer le phénomène afin d'en tirer des informations. Il n'est pas le premier à utiliser ce processus, il existe des manipulations antérieures qu'il importe de connaître. Mais cette connaissance par observation est très souvent insuffisante. Alors il faut se lancer dans une étude de phénomènes et donc expérimenter. Mais l'expérience coûte cher en temps, en moyens matériels à mettre en œuvre. On s'efforcera donc de tirer des expériences faites le maximum d'informations. Pour cela il ne faudra pas faire ces expériences n'importe comment. On cherchera donc à les planifier. C'est cette idée qui est mise en œuvre dans les *plans d'expérience*.

Pour illustrer la démarche on peut prendre une analogie géométrique. Soit une feuille de papier sur laquelle sont dessinés un repère cartésien et une droite. On sait que l'équation de la droite est : $ux + vy + w = 0$, avec des notations évidentes. On veut trouver des valeurs numériques pour u , v , w . Pour cela on prendra deux points sur la droite, on mesurera leurs coordonnées et un calcul simple permettra de répondre à la question. Comment choisir ces deux points ? Compte tenu des erreurs de mesure, il est évident que l'on a intérêt à prendre les points de telle façon que leur distance soit la plus grande possible et non deux proches l'un de l'autre. On a planifié l'expérience.

Depuis longtemps les scientifiques ont pratiqué des expériences destinées à étudier un phénomène. Dès le moyen-âge, Nicolas Oresme en parle. On a voulu ensuite faire des expériences qui puissent être comparées. Ainsi, en 1627, Francis Bacon fait macérer des grains de blé dans neuf concoctions différentes afin d'étudier leur effet sur la rapidité de germination. Arthur Young (1746-1820) systématise le procédé, répète les expériences afin de prendre en compte leur variabilité. Il s'intéresse surtout à l'agronomie et regrette que dans la plupart des traités de son temps les préjugés,

les idées a priori ne soient que trop rarement mis à l'épreuve dans une expérimentation rationnellement menée.

Il reviendra à Sir Ronald Fisher (1890-1962), un des fondateurs de la statistique inductive moderne, d'introduire la notion de plan d'expérience. Il travaille à partir de 1919 dans une station expérimentale agricole à Rothamsted en Angleterre. Il cherche à augmenter les rendements agricoles en combinant type d'engrais, variétés de traitement, méthodes culturales, composition des sols. Il est amené à faire de nombreuses expériences. Compte tenu de la précision qu'il veut obtenir, il veut en réduire le nombre. En effet, il est physiquement impossible, vu le temps que cela prend, de réaliser beaucoup d'expériences. Il faut donc qu'elles soient organisées de telle façon que l'on puisse tirer de celles qui sont faites les informations utiles. Cela l'amène, en particulier, à utiliser des structures mathématiques de nature arithmétique, géométrique, combinatoire, étudiées depuis l'antiquité, cf. l'œuvre de Diophante d'Alexandrie, jusqu'au XIX^{ème} siècle, et qui jusque là figuraient dans la rubrique des récréations mathématiques. La plus célèbre est celle des carrés latins.

De l'agronomie la technique des plans d'expérience s'étendra à d'autres disciplines dont la physique, la chimie. Mais son utilisation nécessitait des connaissances de statistique, de mathématiques relativement peu répandues. La démarche qualité totale vise à atteindre des objectifs en matière de qualité, coût et délai. Il faut donc rechercher le meilleur rapport : avantages tirés de l'information/coût de l'information. Génèchi Taguchi reprendra dans le cadre de la démarche qualité totale la notion de plan d'expérience afin d'aider à la maîtrise de cette qualité dès la conception des produits. A partir de 1975, il vulgarisera la méthode, publiera des tables pratiquement utilisables, mettra au point de nouvelles procédures. Les chimistes, confrontés à des phénomènes mettant en jeu un grand nombre de facteurs, seront des utilisateurs intensifs de la méthode Taguchi et perfectionneront l'outil plan d'expérience.

II. Modélisation et plan d'expérience

1. Le modèle

Planifier les expériences suppose que l'on définisse clairement ce qu'on cherche. Sur le phénomène étudié des facteurs maîtrisables appelés *causes* agissent. Dans l'exemple du paragraphe précédent quatre facteurs ont été identifiés : le type de gazogène, l'essence de bois, le degré d'humidité, la dimension des plaquettes de bois. On peut agir sur ces facteurs, les faire varier. Ensuite, à des états donnés des facteurs correspond une réponse ; cette réponse est quantifiable. Ici il s'agit du rendement en gaz. Dans certains cas la réponse peut être multidimensionnelle. Il en serait ainsi si en plus on s'intéressait à la composition du gaz. La réponse étant fonction des facteurs, si k facteurs ont été identifiés, on peut poser a priori : $y = f(x_1, x_2, \Lambda, x_k)$ avec $y \in \mathbf{R}$ (ou \mathbf{R}^m) est la réponse quantifiable et x_j représente l'état du facteur j , $j = 1, 2, \Lambda, k$. On a $x_j \in E_j$, E_j est l'ensemble des états possibles du facteur j . E_j peut être un ensemble fini : les types de gazogène utilisables, les essences utilisées ; ou un nombre réel : degré d'humidité, volume de la plaquette utilisée. Très souvent, quand une variable représentant un facteur est quantifiable, on se contente de quelques états. Par exemple : petite plaquette, grande plaquette.

Il est bien évident que si l'on répète plusieurs fois une expérience, l'état des facteurs maîtrisés étant inchangé, on obtiendra des réponses différentes. En effet, il y a une fluctuation due aux erreurs de mesure. Plus fondamentalement agissent sur le phénomène des facteurs non pris en compte : la météo, l'habileté des ouvriers etc., dans l'exemple choisi. On sera donc amené à poser :

$$y = f(x_1, x_2, \Lambda, x_k) + \varepsilon$$

On supposera l'expérience menée de telle façon que le "bruit" ε puisse être considéré comme une variable aléatoire. Cette dernière sera de moyenne nulle, d'écart type σ . On fera l'hypothèse qu'elle est gaussienne.

Dans la formulation précédente f est une application de $E_1 \times E_2 \times \Lambda \times E_k$ dans \mathbf{R} ou \mathbf{R}^m qui décrit l'influence des facteurs sur la réponse quantifiable. L'objectif est de caractériser cette fonction f . Dans la plupart des situations réelles, elle ne peut être n'importe quoi. On a déjà une certaine connaissance sur f qui peut se traduire ainsi :

$$y - \varepsilon = f(x_1, x_2, \Lambda, x_k) = g(x_1, x_2, \Lambda, x_k; \lambda_1, \lambda_2, \Lambda, \lambda_p)$$

où g est une fonction connue dépendant des paramètres $\lambda_1, \lambda_2, \Lambda, \lambda_p$. Les expériences faites serviront à donner des valeurs numériques à ces paramètres et si possible à caractériser la précision des mesures faites. On verra ci-dessous des exemples illustrant cette procédure. L'écriture avec la fonction g et l'erreur ε constitue une modélisation de la réalité. Au passage donnons l'illustration suivante : le rendement y d'une réaction chimique industrielle dépend de quatre facteurs : la durée de la réaction : x_1 , la concentration du réactif : x_2 , le ph : x_3 , la température : x_4 . On peut, par exemple postuler le modèle suivant :

$$y = \lambda_0 + \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 + \lambda_4 x_4 + \varepsilon$$

2. L'estimation des paramètres

Pour donner aux paramètres inconnus des valeurs numériques, on fait n expériences. A l'expérience i , $1 \leq i \leq n$, on donne au facteur j la valeur x_{ji} , $x_{ji} \in E_j$. La réponse y_i pourra s'écrire :

$$y_i = g(x_{1i}, x_{2i}, \Lambda, x_{ki}; \lambda_1, \lambda_2, \Lambda, \lambda_p) + \varepsilon_i$$

Toutes les variables aléatoires $\varepsilon_1, \varepsilon_2, \Lambda, \varepsilon_n$ sont supposées indépendantes et de même loi. La valeur attribuée au paramètre λ_l , $l = 1, \Lambda, p$, sera une fonction $\bar{\lambda}_l$ des y_i et des x_{ji} . $\bar{\lambda}_l$ est une variable aléatoire, les valeurs y_i dépendent des ε_i . $\bar{\lambda}_l$ sera appelé *l'estimateur* et sa valeur numérique prise pour les expériences faites *l'estimation*. Quand cela est possible, on cherchera la précision de cette estimation. On peut chercher un intervalle de confiance $[\lambda_{l,L}; \lambda_{l,U}]$ tel que $P[\lambda_{l,L} \leq \lambda_l \leq \lambda_{l,U}] = 1 - \alpha$, α étant un nombre petit fixé d'avance. On prend en général $\alpha = 0,05$, plus rarement $\alpha = 0,01$ ou $\alpha = 0,1$. $\lambda_{l,L}$ et $\lambda_{l,U}$ sont des variables aléatoires dépendant des observations et de σ^2 , variance commune des ε_i . σ est en général inconnu. Il faut l'estimer à partir des observations.

Afin de réaliser ces estimations, il faut un certain nombre d'expériences. Si n , le nombre d'expériences, est inférieur à p , nombre de paramètres, le problème est insoluble. Si $n = p$, on peut estimer les λ_l , mais l'erreur caractérisée par σ reste inconnue et il n'est pas possible de donner un intervalle de confiance. Cela devient possible si $n > p$ et la précision de la mesure sera d'autant plus grande que $n - p$ est grand.

Remarque : Il est des cas où on se pose la question suivante : tel facteur intervient-il dans le phénomène ? Reprenons l'exemple de la réaction chimique industrielle. Si la température n'intervient pas alors $\lambda_4 = 0$. Au vu des expériences, on cherchera alors à tester l'hypothèse suivante $\lambda_4 = 0$ contre l'alternative $\lambda_4 \neq 0$.

3. Validité de la procédure

La procédure générale décrite n'a de sens que si bien entendu le modèle mathématique représenté par la fonction $g(x_1, x_2, \Lambda, x_k; \lambda_1, \lambda_2, \Lambda, \lambda_p)$ où g est connue et $\lambda_1, \lambda_2, \Lambda, \lambda_p$ sont des paramètres à estimer, décrit à peu près la réalité. Tous les modèles hélas sont faux rigoureusement parlant. La réalité ne peut être enfermée dans des schémas si simples. L'essentiel est que la distorsion entre le modèle et la réalité soit, dans les créneaux étudiés, inférieure à l'erreur aléatoire.

Il est essentiel que cette erreur ait les propriétés qui lui sont attribuées : elle est aléatoire, si on considère les n expériences, elles sont indépendantes et de même loi. Cela suppose une certaine façon de conduire les expériences. Illustrons cela par un exemple simple. Un agronome veut connaître l'efficacité d'engrais dans une région donnée. Il doit comparer quatre type d'engrais. Il dispose pour cela de vingt parcelles. A chaque type d'engrais seront affectées cinq parcelles. Mais comment les choisir ? Ces parcelles sont de fertilité inégale. Le facteur fertilité n'est pas pris en compte ; il doit être intégré à l'erreur aléatoire. On fera une erreur en attribuant à un type donné d'engrais les parcelles les plus fertiles. La seule façon de procéder qui évite ces erreurs systématiques appelées *biais* est de tirer au hasard les cinq parcelles attribuées à un engrais spécifié. Cette manière de procéder porte un nom, c'est la *randomisation des expériences*. Random est un nom anglais signifiant aléatoire et qui vient lui-même d'un vieux mot français qui a donné randonnée en français moderne. En chimie on randomise l'attribution d'un numéro d'expérience à une combinaison de facteurs donnée.

III. Le choix des expériences

Une fois le modèle choisi, il faut déterminer quelles sont les combinaisons des facteurs qui feront l'objet d'expériences. Le choix doit être fait pour obtenir des expériences faites la meilleure information possible. L'exemple suivant montre quelles peuvent être les conséquences, en terme de précision des estimateurs, d'une mauvaise planification des expériences.

Soit à étudier un phénomène pour lequel on connaît l'équation de réponse théorique. On la cache et on demande à deux laboratoires d'estimer les effets des 3 facteurs x_i sur la réponse y , en leur laissant le choix du nombre d'essais et de leur structure.

Les variables varient de -1 (valeur minimale codée) à $+1$ (valeur maximale codée) et l'équation de réponse théorique a pour expression :

$$y = 28 + 1,4x_1 + 2,1x_2 + 3,5x_3$$

Les résultats de mesure sont entachés d'une erreur aléatoire, de moyenne nulle et d'écart type égal à 1,414. Ce bruit va avoir une influence très perturbatrice sur les estimations des coefficients, ou relativement négligeable selon la démarche adoptée. C'est ce que montre le tableau suivant, qui indique les résultats des deux laboratoires :

Labo	Nombre de mesures	Constante au centre	coeff. de x_1	coeff. de x_2	coeff. de x_3
1	10	28,55	5,79	-3,86	4,47
2	4	28,05	1,50	2,65	3,60
Valeurs théoriques		28	1,4	2,1	3,5

On voit immédiatement que la *qualité* de l'information n'a rien à voir avec la *quantité* de l'information : le premier laboratoire qui a dépensé le plus d'énergie, trouve des résultats en opposition avec la réalité, tandis que le deuxième, qui s'est contenté du minimum possible d'essais est parvenu à une estimation proche de ce qu'il fallait trouver.

La raison tient au fait que le premier laboratoire a choisi d'augmenter progressivement les trois variables à la fois, selon une méthode « habituelle », qui est valable pour résoudre un problème ponctuel (ici : augmenter la valeur numérique de la réponse) mais qui ne convient pas pour distinguer les effets de chacune des trois variables. La colonne erreur indique la différence entre le constat expérimental et la réponse théorique donnée par l'équation : $y = 28 + 1,4x_1 + 2,1x_2 + 3,5x_3$.

Labo	x_1	x_2	x_3	réponse	erreur
1	-1,0000	-1,0000	-0,8571	22,3	0,8
	-0,5714	-0,4285	-1,0000	22,3	-0,8
	-0,1429	0,0000	0,2857	29,2	0,4
	-0,1429	-0,1429	0,0000	27,0	-0,5
	0,0000	0,1429	0,0000	28,5	0,2
	0,0000	0,0000	0,1429	30,4	1,9
	0,1429	0,0000	0,2857	31,1	1,9
	0,2857	0,4286	0,4286	31,4	0,6
	0,7143	0,8571	1,0000	32,8	-1,5
	1,0000	1,0000	0,8571	34,0	-0,5

Le deuxième laboratoire a fait varier de façon indépendante les trois variables. Pour chaque variable on a choisi les niveaux -1 et 1 . Chaque niveau d'une variable est associé une fois et une seule à un niveau d'une autre variable et cela pour tous les couples de niveaux. On reviendra ultérieurement sur cette propriété fondamentale.

Labo	x_1	x_2	x_3	réponse	erreur
2	-1	$+1$	-1	25,6	0,4
	-1	-1	$+1$	27,5	$-0,5$
	$+1$	$+1$	$+1$	35,8	0,8
	$+1$	-1	-1	23,3	$-0,5$

Pour obtenir des estimations de la constante au centre et des coefficients de chaque variable à partir des constats expérimentaux appelés réponses dans les tableaux précédents, les laboratoires ont utilisé des techniques statistiques classiques, implantées dans les logiciels, que nous ne détaillerons pas.

IV. Un cas particulier important : les facteurs à 2 niveaux

1. Facteurs à deux niveaux

Dans l'étude qui va suivre, on va se restreindre au cas où les facteurs étudiés ne peuvent prendre que deux valeurs. On a donc : $Card E_1 = Card E_2 = \Lambda = Card E_k = 2$. Pour simplifier les écritures on se contentera de prendre $k = 2$ ou $k = 3$. Le cas où les facteurs étudiés sont à deux niveaux est plus fréquent qu'il n'y paraît. Quand on sait à priori qu'un facteur quantitatif a , sur la variable étudiée, un effet se traduisant par une fonction monotone, compte tenu de l'objectif poursuivi : une optimisation, il suffit de l'étudier pour la plus petite valeur possible et la plus grande.

Pour présenter le phénomène on utilisera les notations suivantes dues à Yates : les deux niveaux de chaque facteur seront cotés -1 et $+1$. On a donc $E_j = \{-1, 1\}$. Ces notations sont très pratiques mais malheureusement elles ne peuvent s'étendre au cas où les facteurs ont plus de deux niveaux. Aussi, en vue de cette généralisation, on pose souvent $E_j = \{1, 2\}$.

2. Cas de deux facteurs

Si on reprend la modélisation précédente on aura :

$$y = f(x_1, x_2) + \varepsilon \text{ avec } x_1 = -1 \text{ ou } 1 \text{ et } x_2 = -1 \text{ ou } 1$$

Cette notation étant peu pratique, on posera $f(x_1, x_2) = \mu_{i,j}$ quand x_1 prend la valeur i et x_2 la valeur j . La valeur observée sera alors notée y_{ij} . Éventuellement on introduit un troisième indice quand l'expérience avec $x_1 = i$ et $x_2 = j$ est répétée plusieurs fois. Dans la pratique courante on utilise la même notation y_{ij} pour y_{ij} et $\mu_{i,j}$. Cette dernière valeur est appelée valeur théorique de y_{ij} . C'est le contexte qui permet de savoir de quoi on parle.

On va maintenant, à partir des valeurs $\mu_{i,j}$, définir un effet global, un effet pour chaque facteur. Il y aura un « résidu » qui sera analysé après.

a) L'effet global est l'effet de chaque facteur

L'effet global est naturellement la moyenne de toutes les valeurs possibles ; elle sera notée $\mu_{\bullet\bullet}$:

$$\mu_{\bullet\bullet} = \frac{1}{4}(\mu_{-1,-1} + \mu_{-1,1} + \mu_{1,-1} + \mu_{1,1}).$$

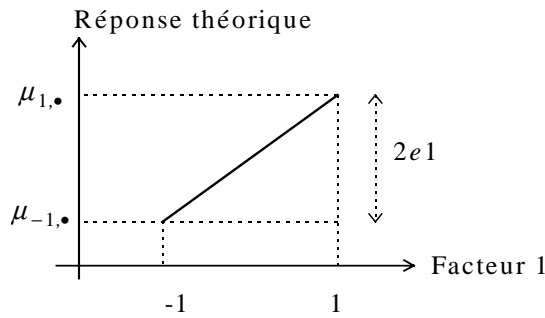
Tout aussi naturellement on peut définir l'effet du premier facteur par la différence des moyennes quand il est au niveau 1 et quand il est au niveau -1 . Pour des raisons de généralisation ultérieure on préfère prendre la demi-différence. Appelons e_1 cet effet. On aura :

$$e_1 = \frac{1}{2} \left[\frac{\mu_{1,-1} + \mu_{1,1}}{2} - \frac{\mu_{-1,-1} + \mu_{-1,1}}{2} \right]$$

Si on note $\mu_{i,\bullet}$ la moyenne des $\mu_{i,j}$ quand le premier facteur est à l'état i , un calcul simple montre que $e_1 = \mu_{1,\bullet} - \mu_{\bullet\bullet} = -(\mu_{-1,\bullet} - \mu_{\bullet\bullet})$, on définit de même l'effet du deuxième facteur : $e_2 = \mu_{\bullet,+1} - \mu_{\bullet\bullet} = -(\mu_{\bullet,-1} - \mu_{\bullet\bullet})$, on dit aussi que l'effet du facteur 1 quand il est au niveau 1 est mesuré par $\mu_{1,\bullet} - \mu_{\bullet\bullet}$ et par $\mu_{-1,\bullet} - \mu_{\bullet\bullet}$ quand il est au niveau -1 . Les deux valeurs sont ici opposées. Ce langage a le mérite de se généraliser au cas où il y a plus de deux niveaux par facteur. On pose $e_1(1) = \mu_{1,\bullet} - \mu_{\bullet\bullet}$ et $e_1(-1) = \mu_{-1,\bullet} - \mu_{\bullet\bullet}$. On a $e_1(1) + e_1(-1) = 0$.

Si le facteur 1 avait trois niveaux notés 1, 2, 3 on définirait : $e_1(1) = \mu_{1,\bullet} - \mu_{\bullet\bullet}$, $e_1(2) = \mu_{2,\bullet} - \mu_{\bullet\bullet}$, $e_1(3) = \mu_{3,\bullet} - \mu_{\bullet\bullet}$ et on aurait : $e_1(1) + e_1(2) + e_1(3) = 0$

Les effets des facteurs peuvent se représenter graphiquement :



b) L'interaction

Isoler l'effet de chaque facteur ne rend pas compte toujours de l'ensemble du phénomène. Il arrive que l'un des facteurs renforce l'effet de l'autre. Cela s'appelle une interaction. On illustre le phénomène par les graphiques suivants :

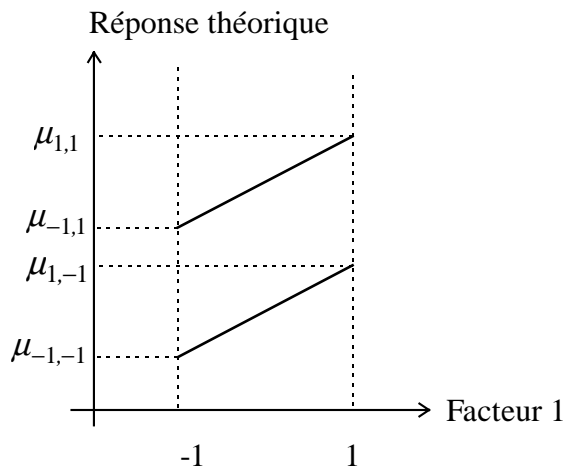


Fig. 1 Sans interaction

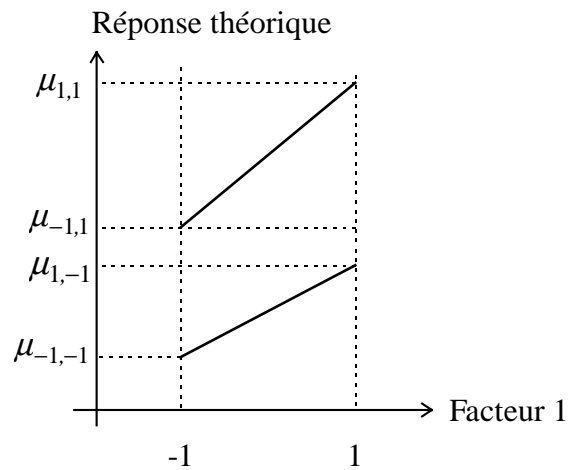


Fig. 2 Avec interaction

Dans chacune des deux figures le segment du bas représente les variations de la réponse théorique quand on passe du niveau -1 au niveau 1 pour le premier facteur, le deuxième étant au niveau -1 . De même le segment du haut représente les variations de la réponse théorique quand on passe du niveau -1 au niveau 1 pour le premier facteur, le deuxième étant alors au niveau $+1$. Dans la première figure les segments sont parallèles. Cela signifie que l'effet du facteur 1 est le même quel que soit l'état du facteur 2 . L'accroissement de la réponse est le même quel que soit le deuxième facteur. On a :

$$\mu_{1,1} - \mu_{-1,1} = \mu_{1,-1} - \mu_{-1,-1}$$

Dans la deuxième figure, le segment du haut a une pente plus forte que celui du bas et on a :

$$\mu_{1,1} - \mu_{-1,1} > \mu_{1,-1} - \mu_{-1,-1}$$

Cela veut dire que l'effet du premier facteur est plus fort quand le deuxième facteur est au niveau 1 que quand il est au niveau -1. Le deuxième facteur renforce l'action du premier ; il y a interaction. On se sert de la différence entre les deux accroissements pour mesurer l'interaction :

$$I_{1,2} = \frac{1}{4} \left((\mu_{1,1} - \mu_{-1,1}) - (\mu_{1,-1} - \mu_{-1,-1}) \right)$$

Pour illustrer le phénomène nous avons dissymétrisé le problème : le facteur 1 variait, le facteur 2 restait constant. L'examen de la formule montre que l'on aurait obtenu le même résultat en faisant varier le facteur 2, le facteur 1 restant constant. On a donc $I_{1,2} = I_{2,1}$. A titre d'exercice on pourra vérifier les formules suivantes :

$$I_{1,2} = \mu_{1,1} - e_1 - e_2 - \mu_{\bullet\bullet}$$

(c'est à dire la réponse quand les deux facteurs sont au niveau 1 diminuée des effets de chaque facteur et de l'effet global)

$$\text{et } I_{1,2} = \mu_{1,1} - \mu_{1,\bullet} - \mu_{\bullet,1} + \mu_{\bullet\bullet}$$

Ces formules sont utiles quand on veut généraliser au cas où il y a plus de deux niveaux par facteur. On a, jusqu'à présent, caractérisé l'interaction des deux facteurs par un seul nombre $I_{1,2}$. On va maintenant définir l'effet de l'interaction quand le facteur 1 prend l'état i et le facteur 2 l'état j ; il sera noté $I_{1,2}(i, j)$. Ce sera ce qui reste quand on fait abstraction de l'effet global et de l'effet de chaque facteur :

$$I_{1,2}(i, j) = \mu_{i,j} - e_1(i) - e_2(j) - \mu_{\bullet\bullet} = \mu_{i,j} - \mu_{i,\bullet} - \mu_{\bullet,j} + \mu_{\bullet\bullet}$$

On remarquera, dans le cas qui nous intéresse, que :

$$I_{1,2}(1, 1) + I_{1,2}(1, -1) = 0$$

$$I_{1,2}(-1, 1) + I_{1,2}(-1, -1) = 0$$

$$I_{1,2}(-1, 1) + I_{1,2}(1, 1) = 0$$

$$I_{1,2}(-1, -1) + I_{1,2}(1, -1) = 0$$

Un seul nombre suffit donc à caractériser les quatre valeurs de l'interaction.

On remarque que l'on peut écrire :

$$\mu_{i,j} = \mu_{\bullet\bullet} + (\mu_{i,\bullet} - \mu_{\bullet\bullet}) + (\mu_{\bullet,j} - \mu_{\bullet\bullet}) + (\mu_{i,j} - \mu_{i,\bullet} - \mu_{\bullet,j} + \mu_{\bullet\bullet})$$

Cette écriture est triviale mais elle permet l'interprétation suivante :

$\mu_{\bullet\bullet} = \mu$ est l'effet global

$\mu_{i,\bullet} - \mu_{\bullet\bullet} = e1(i)$ est l'effet du premier facteur au niveau i

$\mu_{\bullet,j} - \mu_{\bullet\bullet} = e2(j)$ est l'effet du deuxième facteur au niveau j

$\mu_{i,j} - \mu_{i,\bullet} - \mu_{\bullet,j} + \mu_{\bullet\bullet} = I1,2(i, j)$ est l'interaction des deux facteurs aux niveaux respectifs i et j .

$$\begin{aligned} \text{On a de plus : } \sum_i e1(i) &= \sum_j e2(j) = 0 \\ \sum_j I1,2(i, j) &= 0 \text{ quel que soit } i \\ \sum_i I1,2(i, j) &= 0 \text{ quel que soit } j. \end{aligned}$$

Si on retourne aux notations avec les x_j , x_1 représente le premier facteur et prend les valeurs -1 ou 1 avec les notations de Yates, de même pour x_2 représentant le deuxième facteur, les égalités précédentes permettent d'écrire :

$$\mu_{ij} = \mu + e1 \cdot x_1 + e2 \cdot x_2 + I1,2 \cdot x_1 \cdot x_2$$

Le modèle devient :

$$y = \mu + e1 \cdot x_1 + e2 \cdot x_2 + I1,2 \cdot x_1 \cdot x_2 + \varepsilon$$

Dans certains cas, on suppose que l'interaction est nulle ; on a alors le modèle :

$$y = \mu + e1 \cdot x_1 + e2 \cdot x_2 + \varepsilon$$

c) Notion de degré de liberté

On appelle action les différentes composantes qui ont été définies précédemment,

on distingue donc :

- l'action globale dont l'effet est μ
- l'action 1 du premier facteur
- l'action 2 du deuxième facteur
- l'action (1, 2) résultant de l'interaction des deux facteurs.

On appelle degré de liberté le nombre de paramètres nécessaires à la description de chaque action. Ici toutes les actions ont le même nombre de degré de liberté : 1.

3. Cas de trois facteurs

On suppose maintenant que l'on a trois facteurs, chacun pouvant prendre deux niveaux. Selon les cas on prend la notation de Yates : $E_1 = E_2 = E_3 = \{-1, +1\}$ ou la notation pouvant s'étendre à des cas plus généraux : $E_1 = E_2 = E_3 = \{1, 2\}$. Les facteurs seront appelés A, B, C . Si A est au niveau i, B au niveau j, C au niveau k , la réponse théorique sera notée : $\mu_{i,j,k}$.

Comme précédemment on pose :

$$\mu_{\dots} = \frac{1}{8} \sum_i \sum_j \sum_k \mu_{i,j,k}, \text{ moyenne générale des } 2^3 = 8 \text{ valeurs possibles pour les } \mu_{i,j,k}.$$

$$\mu_{i\bullet\bullet} = \frac{1}{4} \sum_j \sum_k \mu_{i,j,k}, \text{ moyenne des } \mu_{i,j,k}, i \text{ étant fixé.}$$

De même :

$$\mu_{\bullet j\bullet} = \frac{1}{4} \sum_i \sum_k \mu_{i,j,k}$$

$$\mu_{\bullet\bullet k} = \frac{1}{4} \sum_i \sum_j \mu_{i,j,k}$$

$$\mu_{ij\bullet} = \frac{1}{2} \sum_k \mu_{i,j,k}, \text{ moyenne des } \mu_{i,j,k}, i \text{ et } j \text{ étant fixés.}$$

$$\mu_{i\bullet k} = \frac{1}{2} \sum_j \mu_{i,j,k}$$

$$\mu_{\bullet jk} = \frac{1}{2} \sum_i \mu_{i,j,k}$$

On peut étendre ce qui a été fait pour deux facteurs.

On définit :

- l'effet global mesuré par μ_{\dots}
- l'effet du facteur A : eA , mesuré en i par $eA(i) = \mu_{i\bullet\bullet} - \mu_{\dots}$; on a $eA(-1) = -eA(1) = -eA$,
- l'effet du facteur B : eB , mesuré en j par $eB(j) = \mu_{\bullet j\bullet} - \mu_{\dots}$; on a $eB(-1) = -eB(1) = -eB$,
- l'effet du facteur C : eC , mesuré en k par $eC(k) = \mu_{\bullet\bullet k} - \mu_{\dots}$; on a $eC(-1) = -eC(1) = -eC$,
- l'interaction des facteurs A et B :
 IAB mesurée en i, j par $IAB(i, j) = \mu_{ij\bullet} - \mu_{i\bullet\bullet} - \mu_{\bullet j\bullet} + \mu_{\dots}$;
on a $IAB(1, 1) = -IAB(-1, 1) = -IAB(1, -1) = IAB(-1, -1) = IAB$
- l'interaction des facteurs B et C ,

- l'interaction des facteurs A et C .

On laisse au lecteur le soin d'écrire leurs valeurs.

a) L'interaction d'ordre 3

Dans le cas des deux variables, on avait vu que l'effet global et les effets de chaque facteur n'épuisaient pas la description. On a alors introduit l'interaction entre les deux facteurs. Il en est de même ici. Si on fixe le facteur C à une valeur donnée, on a alors un système à deux facteurs et on peut calculer leur interaction. La demi-différence entre les deux interactions calculées l'une quand C est au niveau 1, l'autre quand C est au niveau -1 , mesure l'influence de C sur l'interaction AB . Cela définit l'interaction $IABC$ des trois facteurs. Elle est cohérente car le rôle particulier joué par C n'est qu'apparent. On obtiendrait le même résultat en mesurant les variations de IAC quand B passe du niveau -1 au niveau 1 ou de IBC quand A passe du niveau -1 au niveau 1. Un calcul simple montre que l'on peut définir l'effet de l'interaction d'ordre 3 quand les facteurs sont aux niveaux respectifs i, j, k par :

$$IABC(i, j, k) = \mu_{i,j,k} - \mu_{\dots} - eA(i) - eB(j) - eC(k) - IAB(i, j) - IAC(i, k) - IBC(j, k).$$

Si on considère $IABC(1, 1, 1)$, on a :

$$IABC(1, 1, -1) = IABC(1, -1, 1) = IABC(-1, 1, 1) = -IABC(1, 1, 1)$$

$$IABC(1, -1, -1) = IABC(-1, 1, -1) = IABC(-1, -1, 1) = IABC(1, 1, 1)$$

$$IABC(-1, -1, -1) = -IABC.$$

On change de signe à chaque fois qu'un niveau d'un des facteurs varie. L'interaction d'ordre 3 est décrite par un paramètre ; son nombre de degré de liberté est 1.

b) Les modèles dans le cas de trois facteurs

Le modèle général peut alors s'écrire :

$$y_{i,j,k} = \mu + eA(i) + eB(j) + eC(k) + IAB(i, j) + IAC(i, k) + IBC(j, k) + IABC(i, j, k) + \varepsilon$$

Compte tenu des définitions ci-dessus cette écriture est une autre façon d'écrire :

$$y_{i,j,k} = \mu_{i,j,k} + \varepsilon$$

Si x_i, x_j, x_k représentent les niveaux respectifs des facteurs A, B, C , avec les notations de Yates, on peut également écrire :

$$y_{i,j,k} = \mu + eA \cdot x_i + eB \cdot x_j + eC \cdot x_k + IAB \cdot x_i \cdot x_j + IAC \cdot x_i \cdot x_k + IBC \cdot x_j \cdot x_k + IABC \cdot x_i \cdot x_j \cdot x_k + \varepsilon$$

Il y a alors huit paramètres pour décrire le système : $\mu, eA, eB, eC, IAB, IAC, IBC, IABC$.

Dans beaucoup de phénomènes étudiés dans l'industrie chimique on peut faire l'hypothèse que l'interaction d'ordre 3 est nulle : $IABC = 0$. Il reste alors sept paramètres pour décrire le système. D'autres simplifications sont possibles. Il y a de nombreux cas où l'on sait a priori que l'une ou l'autre des interactions est nulle.

c) Exemple numérique

On s'intéresse à la capabilité d'une machine mécanique. Celle-ci fabrique en grande série des pièces dont le diamètre doit s'écarter le moins possible de la norme exigée. La capabilité est alors définie comme l'écart type de la série des diamètres d'un grand nombre de pièces exprimé en centième de millimètre. Cette capabilité sera considérée comme la réponse du système.

Pour faire fonctionner cette machine on a le choix entre :

- deux types de plaquettes, facteur noté A ,
- deux vitesses de coupe, facteur noté B
- deux types de lubrifiant, facteur noté C

Les capabilités théoriques sont supposées connues et, pour chaque combinaison des facteurs, données par le tableau suivant :

			B		
			-1	1	
A	-1	C	-1	2,9	1,7
		C	1	1,9	0,8
	1	C	-1	3,6	2,4
		C	1	4,6	3,4

Calculons les différents effets :

$$\text{effet global : } \mu = \frac{1}{8}(2,9 + 1,9 + 3,6 + 4,6 + 1,7 + 0,8 + 2,4 + 3,4) = \frac{1}{8} \cdot 21,3$$

$$\text{effet } A : eA = \frac{1}{4}(3,6 + 4,6 + 2,4 + 3,4) - \mu = \frac{1}{8} \cdot 6,7$$

$$\text{effet } B : eB = \frac{1}{4}(3,6 + 0,8 + 2,4 + 3,4) - \mu = \frac{1}{8} \cdot 4,7$$

$$\text{effet } C : eC = \frac{1}{4}(1,9 + 4,6 + 0,8 + 3,4) - \mu = \frac{1}{8} \cdot 0,1$$

$$\text{Interaction } AB : IAB = \frac{1}{2}(2,4 + 3,4) - \frac{1}{4}(3,6 + 4,6 + 2,4 + 3,4) - \frac{1}{4}(1,7 + 0,8 + 2,4 + 3,4) + \mu = \frac{1}{8} \cdot 0,1$$

$$\text{Interaction } AC : IAC = \frac{1}{2}(1,9 + 4,6) - \frac{1}{4}(3,6 + 4,6 + 2,4 + 3,4) - \frac{1}{4}(1,9 + 4,6 + 0,8 + 3,4) + \mu = \frac{1}{8} \cdot 3,9$$

$$\text{Interaction } BC : IBC = \frac{1}{2}(0,8 + 3,4) - \frac{1}{4}(1,7 + 0,8 + 2,4 + 3,4) - \frac{1}{4}(1,9 + 4,6 + 0,8 + 3,4) + \mu = \frac{1}{8} \cdot 0,1$$

$$\text{Interaction } ABC : IABC = 3,4 - IAB - IAC - IBC - eA - eB - eC - \mu = \frac{1}{8} \cdot 0,1$$

$$\begin{array}{llll} \text{En effectuant : } \mu = 2,7 & eA = 0,8 & IAB = 0 & IABC = 0 \\ & eB = -0,6 & IAC = 0,5 & \\ & eC = 0 & IBC = 0 & \end{array}$$

On remarque que l'interaction d'ordre 3 et les interactions de A et B , de B et C sont nulles. Le modèle serait donc lors d'une expérimentation :

$$y = 2,7 + 0,8 \cdot x_1 - 0,6 \cdot x_2 + 0,5 \cdot x_1 \cdot x_3 + \varepsilon$$

Notons que l'effet de C est nul mais que C joue un rôle par l'intermédiaire de l'interaction de A et de C .

V. Un premier plan d'expérience : le plan factoriel complet

1. Définition et notations

Reprenons la problématique du paragraphe II. Nous avons une réponse quantifiable y sous la dépendance de k facteurs. Le facteur j est décrit par la variable x_j prenant ses valeurs dans un ensemble E_j , j variant de 1 à k . On a établi un modèle et on sait que :

$$y = g(x_1, x_2, \dots, x_k; \lambda_1, \lambda_2, \dots, \lambda_p) + \varepsilon$$

où g est une fonction connue, $\lambda_1, \lambda_2, \dots, \lambda_p$ des paramètres à déterminer, ε une perturbation aléatoire suivant une loi de Gauss de moyenne 0 et d'écart type σ .

Supposons que les E_j soient des ensembles finis avec $\text{Card } E_j = n_j$. L'idée première pour une expérimentation rationnelle consiste à essayer toutes les combinaisons possibles de tous les facteurs. On fera ainsi $N = n_1 \times n_2 \times \dots \times n_k$ expériences. Si $p \leq N$, il sera possible, en général, d'estimer $\lambda_1, \lambda_2, \dots, \lambda_p$; si $p < N$, on pourra en outre estimer σ .

Dans le cas étudié précédemment nous avons $n_j = 2$. Il faudra faire alors 2^k expériences. Quand $k = 3$, cela en fait huit. On va les numéroter de 1 à 8. Le tableau suivant donne les niveaux des quatre facteurs A, B, C affectés à chaque expérience.

n°	A	B	C
1	-1	-1	-1
2	-1	-1	1
3	-1	1	-1
4	-1	1	1
5	1	-1	-1
6	1	-1	1
7	1	1	-1
8	1	1	1

a) Nouvelles notations pour le cas 2^k

Afin de simplifier les notations, on appellera du même nom le facteur et la variable qui le représente. Ainsi A, B, C prend les valeurs -1 ou 1 et on posera $\mu(A, B, C)$ la réponse théorique (avec les notations précédentes on notait par exemple $\mu_{1,1,1}$ ce que l'on appelle ici $\mu(1,1,1)$). L'analyse précédente nous avait fait distinguer l'effet global que l'on appellera ici $e(1)$, les effets de chaque facteurs $e(A), e(B), e(C)$, les effets des interactions d'ordre 2 notées ici $e(AB), e(BC), e(AC)$ au lieu de IAB, IBC, IAC , les effets de l'interaction d'ordre 3 notée ici $e(ABC)$ au lieu de $IABC$.

On pourra vérifier qu'avec ces nouvelles notations on a :

$$\begin{aligned}
 e(1) &= \frac{1}{8} \sum_{A,B,C} \mu(A, B, C) \\
 e(A) &= \frac{1}{8} \sum_{A,B,C} A \mu(A, B, C) \\
 e(AB) &= \frac{1}{8} \sum_{A,B,C} AB \mu(A, B, C) \\
 e(ABC) &= \frac{1}{8} \sum_{A,B,C} ABC \mu(A, B, C)
 \end{aligned}$$

On peut résumer tout cela en une formule unique :

$$e(A^a B^b C^c) = \frac{1}{8} \sum_{A,B,C} A^a B^b C^c \mu(A, B, C)$$

où les exposants a, b, c prennent les valeurs 0 ou 1. Ainsi $e(A) = e(A^1 B^0 C^0)$.

Le phénomène étudié a été dit sous l'effet de diverses actions : l'action globale, l'action de chaque facteur, l'action de chaque interaction. Dans le cas qui nous occupe : trois facteurs à deux niveaux, chaque action a un degré de liberté ; elle est caractérisée par un seul paramètre. Suivant le

niveau des facteurs d'une expérience particulière, ce paramètre est affecté du signe + ou du signe -. Prenons un exemple pour illustrer cela.

Avec les définitions et les notations du paragraphe IV on a par exemple :

$$\mu(1,1,-1) = \mu + eA(1) + eB(1) + eC(-1) + I AB(1,-1) + I AC(1,-1) + I BC(1,-1) + I ABC(1,1,-1)$$

ce qui peut s'écrire avec les nouvelles notations :

$$\mu(1,1,-1) = \mu + eA + eB - eC + e(AB) - e(AC) - e(BC) - e(ABC).$$

Plus généralement on a :

$$\mu(A, B, C) = \mu + Ae(A) + Be(B) + Ce(C) + AB e(AB) + AC e(AC) + BC e(BC) + ABC e(ABC)$$

On rappelle que l'on a appelé *action* chaque composante de la décomposition ci-dessus :

l'action 1 produit l'effet moyen

l'action A produit l'effet du facteur A

l'action AB produit l'effet de l'interaction AB. De même pour les actions B, C, AC, BC, ABC.

On remarquera que le niveau d'une interaction est le produit des niveaux des facteurs qui la composent. En effectuant ces produits on obtient la matrice suivante donnant les niveaux de chaque action pour une expérience donnée :

n° d'expérience	valeurs théoriques	1	A	B	C	AB	AC	BC	ABC
1	$\mu(-1,-1,-1)$	1	-1	-1	-1	1	1	1	-1
23	$\mu(-1,-1,1)$	1	-1	-1	1	1	-1	-1	1
3	$\mu(-1,1,-1)$	1	-1	1	-1	-1	1	-1	1
4	$\mu(-1,1,1)$	1	-1	1	1	-1	-1	1	-1
5	$\mu(1,-1,-1)$	1	1	-1	-1	-1	-1	1	1
6	$\mu(1,-1,1)$	1	1	-1	1	-1	1	-1	-1
7	$\mu(1,1,-1)$	1	1	1	-1	1	-1	-1	-1
8	$\mu(1,1,1)$	1	1	1	1	1	1	1	1

Dans ce qui suit on illustrera les calculs par l'exemple suivant. Un produit de base utile dans l'industrie des colorants est produit par une réaction où intervient l'acide nitrique NO_3H . Trois facteurs sont pris en compte :

A : le temps pendant lequel dure l'addition d'acide nitrique deux niveaux sont testés 2 heures (noté -1) et 7 heures (noté +1).

B : le temps pendant lequel on agite le mélange, deux niveaux sont testés : une demi-heure (notée -1) et quatre heures (noté +1).

C : le fait ou non que le récipient ayant servi à la réaction soit nettoyé des résidus de la réaction précédente ; on notera -1 l'absence de nettoyage et $+1$ sa présence.

On mesure le pourcentage de matière utile produite. On fait les huit expériences prévues ci-dessus.

On obtient les résultats suivants. $y(A, B, C)$ sera le pourcentage de matière produite moins 80.

n° d'expérience	A	B	C	$y(A, B, C)$
1	-1	-1	-1	7,2
2	-1	-1	1	6,7
3	-1	1	-1	2
4	-1	1	1	3,4
5	1	-1	-1	8,4
6	1	-1	1	9,2
7	1	1	-1	3
8	1	1	1	3,7

Dans le tableau ci-dessus, les $y(A, B, C)$ sont, d'un point de vue mathématique, les valeurs prises par des variables aléatoires.

2. Estimation des paramètres, modèle général

Si on ne fait pas l'hypothèse qu'une action donnée a un effet nul nous avons huit paramètres à estimer : l'action 1, les actions des facteurs A, B, C , les trois actions interaction d'ordre deux et l'action interaction d'ordre trois. Nous avons huit expériences. On sait que $y(A, B, C)$ est la valeur prise par une variable aléatoire gaussienne de moyenne $\mu(A, B, C)$. Donc pour estimer un effet, il suffira dans l'expression donnant l'effet de l'action de remplacer $\mu(A, B, C)$ par $y(A, B, C)$. On a :

$$e(A^a B^b C^c) = \frac{1}{8} \sum_{A,B,C} A^a B^b C^c \mu(A, B, C) \text{ avec } a, b, c \text{ prenant les valeurs } 0 \text{ ou } 1.$$

$$\text{Il en résulte que son estimation sera : } \bar{e}(A^a B^b C^c) = \frac{1}{8} \sum_{A,B,C} A^a B^b C^c y(A, B, C).$$

La matrice donnant les niveaux de chaque action donne les coefficients dont il faut affecter les observations pour obtenir les estimations désirées. Le résultat de la $i^{\text{ème}}$ expérience sera noté aussi y_j . On a aussi par exemple :

$$\bar{e}(1) = \frac{1}{8} (y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8)$$

$$\bar{e}(A) = \frac{1}{8} (-y_1 - y_2 - y_3 - y_4 + y_5 + y_6 + y_7 + y_8)$$

$$\bar{e}(AB) = \frac{1}{8} (y_1 + y_2 - y_3 - y_4 - y_5 - y_6 + y_7 + y_8)$$

Toutes ces estimations sont des valeurs prises par des moyennes de variables aléatoires gaussiennes indépendantes. La moyenne d'un estimateur est la quantité qu'il estime, sa variance est la somme des variances multipliée par $\frac{1}{64}$. Si σ_e est l'écart type d'un quelconque de ces estimateurs

$$\text{on a : } \sigma_e = \frac{\sigma}{\sqrt{8}} = \frac{\sigma}{2\sqrt{2}}.$$

Comme il y avait huit expériences pour huit paramètres, il n'est pas possible d'estimer σ .

On vérifiera que pour l'exemple précédent on obtient les estimations suivantes :

paramètre	estimation
$e(1)$	5,45
$e(A)$	0,625
$e(B)$	-2,425
$e(C)$	0,3
$e(AB)$	-0,3
$e(AC)$	-0,075
$e(BC)$	0,225
$e(ABC)$	-0,25

3. Estimation des paramètres : modèle particulier

Supposons que l'on sache a priori que certaines interactions sont nulles. Par exemple postulons que : $\mu(A, B, C) = e(1) + Ae(A) + Be(B) + Ce(C) + ABe(AB)$. Les interactions AC , BC , ABC sont supposées nulles. Les estimations des paramètres restants seront les mêmes que dans le cas général. Mais on a huit expériences pour cinq paramètres à estimer. Reste trois degrés de liberté qui nous serviront à estimer σ .

Pour chacune des huit expériences on peut comparer la valeur prise par $y : y(A, B, C)$ et la valeur calculée à partir du modèle $\tilde{y}(A, B, C)$ où :

$$\tilde{y}(A, B, C) = \bar{\mu}(A, B, C) = \bar{\mu}(1) + A\bar{\mu}(A) + B\bar{\mu}(B) + C\bar{\mu}(C) + AB\bar{\mu}(AB)$$

Les huit valeurs $y(A, B, C) - \tilde{y}(A, B, C)$ donnent une idée de la fluctuation aléatoire. On montre qu'une bonne estimation de σ , notée $\bar{\sigma}^2$ est donnée par la formule :

$$\bar{\sigma}^2 = \frac{1}{8-5} \sum_{A,B,C} [y(A, B, C) - \tilde{y}(A, B, C)]^2$$

Le dénominateur est déterminé de la façon suivante :

$8 - 5 =$ nombre d'expériences - nombre de paramètres estimés. Si on applique ce schéma à l'exemple pris on a :

n° d'expérience	$y(A, B, C)$	$\tilde{y}(A, B, C)$	$y(A, B, C) - \tilde{y}(A, B, C)$
1	7,2	6,65	0,55
2	6,7	7,25	-0,55
3	2	2,4	-0,4
4	3,4	3	0,4
5	8,4	8,5	-0,1
6	9,2	9,1	0,1
7	3	3,05	-0,05
8	3,7	3,65	0,05

(Dans le cas du modèle général les $y - \tilde{y}$ sont évidemment tous nuls)

d'où $\bar{\alpha} = 0,56$.

On observe que les résidus $y(A, B, C) - \tilde{y}(A, B, C)$ sont, en valeur absolue, inférieurs à un écart type. Cela ne constitue ni un test statistique de validité, ni une preuve, mais reste cohérent avec le modèle.

4. Propriété du plan factoriel : l'orthogonalité

La façon dont est construit le plan factoriel permet de vérifier la propriété suivante :

propriété : *pour tout couple de facteurs, chaque niveau de l'un est associé à chaque niveau de l'autre un même nombre de fois indépendamment du couple de facteur choisi.*

Un plan présentant cette propriété est dit orthogonal. On vérifie que le plan factoriel 2^3 est orthogonal : chaque niveau d'un des facteurs est associé deux fois à un niveau d'un autre facteur. Cette propriété est fondamentale pour la précision des estimations. Un effet est estimé en faisant intervenir tous les résultats des expériences faites. Un estimateur est une variable aléatoire d'écart type $\frac{\sigma}{\sqrt{n}}$, où n est le nombre d'expériences prises en compte pour l'estimation faite. La précision est mesurée par cet écart type. Elle est donc maximum quand n est le plus grand possible.

VI. Un plan orthogonal fractionnaire

1. Problématique

Reprenons le cas où la réponse est fonction de trois facteurs maîtrisables à deux niveaux chacun. Supposons que toutes les interactions soient nulles. Le modèle devient :

$$y = e(1) + Ae(A) + Be(B) + Ce(C) + \varepsilon$$

avec les notations du précédent chapitre. Nous avons quatre paramètres à estimer : $e(1)$, $e(A)$, $e(B)$, $e(C)$. Le plan factoriel impose huit expériences. Cela permet certes l'estimation de l'écart type σ de ε avec une bonne précision. Mais si le coût des expériences est élevé cela est un luxe. Pour estimer quatre paramètres il faut au minimum quatre expériences. On va donc essayer de trouver un plan ne comportant que quatre expériences. Pour des raisons de précision des estimations et de simplicité de celles-ci nous imposerons à ce plan d'être orthogonal.

2. Construction du plan et estimations liées

Considérons la matrice du V 1 qui donne en ligne pour une expérience les niveaux de chaque action :

1	A	B	C	AB	AC	BC	ABC	
1	-1	-1	-1	1	1	1	-1	
1	-1	-1	1	1	-1	-1	1	←
1	-1	1	-1	-1	1	-1	1	←
1	-1	1	1	-1	-1	1	-1	
1	1	-1	-1	-1	-1	1	1	←
1	1	-1	1	-1	1	-1	-1	
1	1	1	-1	1	-1	-1	-1	
1	1	1	1	1	1	1	1	←

Parmi les huit expériences possibles sélectionnons celles qui mettent l'interaction ABC au niveau 1. Puisque cette interaction est nulle, il est inutile de la faire varier pour l'estimer. On fera alors quatre expériences notées de 1 à 4. On obtient le plan d'expérience suivant :

n° d'expérience	1	A	B	C	AB	AC	BC	ABC
1	1	-1	-1	1	1	-1	-1	1
2	1	-1	1	-1	-1	1	-1	1
3	1	1	-1	-1	-1	-1	1	1
4	1	1	1	1	1	1	1	1

Quoique les interactions soient supposées nulles, on a fait figurer les niveaux des actions correspondantes pour une interprétation ultérieure. On peut vérifier que ce nouveau plan est orthogonal : un niveau d'un facteur donné est associé à un niveau d'un autre facteur.

Si on appelle y_i le résultat de l'expérience i , on peut comme précédemment estimer $e(1)$, $e(A)$, $e(B)$, $e(C)$. On a :

$$\bar{e}(1) = \frac{1}{4}(y_1 + y_2 + y_3 + y_4)$$

$$\bar{e}(A) = \frac{1}{4}(-y_1 - y_2 + y_3 + y_4)$$

$$\bar{e}(B) = \frac{1}{4}(-y_1 + y_2 - y_3 + y_4)$$

$$\bar{e}(C) = \frac{1}{4}(y_1 + y_2 - y_3 - y_4)$$

Ces estimations sont les valeurs prises par des variables aléatoires gaussiennes dont la moyenne est ce qu'il faut pour estimer et ils ont tous le même écart type σ_1 avec $\sigma_1 = \frac{\sigma}{2}$. Avec les huit expériences du plan factoriel, l'écart type était $\sigma_0 = \frac{\sigma}{2\sqrt{2}}$. En divisant par deux le nombre d'expériences on a augmenté l'incertitude de l'estimation d'un facteur $\sqrt{2}$.

Les estimations précédentes sont des applications de la formule :

$$\bar{e}(A^a B^b C^c) = \frac{1}{4} \sum_{A,B,C} A^a B^b C^c y(A, B, C) \quad a, b, c \text{ prenant les valeurs } 0 \text{ ou } 1,$$

la somme sur A, B, C étant restreinte aux expériences faites, elle comporte quatre termes au lieu de huit dans le plan factoriel. On notera que c'est l'orthogonalité du plan qui permet la simplicité de la formule.

Il y a quatre expériences et quatre paramètres à estimer. Il ne sera donc pas possible avec ce plan fractionnaire d'estimer σ .

En application reprenons l'exemple de la réaction avec l'acide nitrique. En sélectionnant les expériences 2, 3, 5, 8 qui correspondent au plan fractionnaire et en renumérotant on a :

n° d'expérience	A	B	C	y(ABC)
1	-1	-1	1	6,7
2	-1	1	-1	2
3	1	-1	-1	8,4
4	1	1	1	3,7

On vérifiera les valeurs des estimations :

paramètre	estimation
$e(1)$	5,2
$e(A)$	0,85
$e(B)$	-2,35
$e(C)$	0

Remarque : le plan fractionnaire utilisé ci-dessus n'est pas le seul possible. On aurait pu sélectionner les expériences qui donnent la valeur -1 à l'interaction d'ordre 3 ou bien celle qui donnent une valeur fixée d'avance à l'une quelconque des interactions d'ordre 2. Comme toutes ces interactions sont supposées nulles, il n'est pas besoin de les faire varier pour les estimer. Cela fait huit plans fractionnaires orthogonaux possibles.

3. Les alias ou la confusion des actions

On a supposé que toutes les interactions étaient nulles dans le modèle initial. Que se passe-t-il quand cela est faux et que l'on utilise quand même un plan fractionnaire ?

Examinons le tableau des niveaux des actions donné au paragraphe précédent. On observe que chaque colonne est répétée deux fois. Sont identiques les colonnes suivantes :

1 et ABC

A et BC

B et AC

C et AB

Cela veut dire que quand on veut estimer $e(1)$ on estime en réalité $e(1) + e(ABC)$. On dit que les actions action globale et interactions d'ordre 3 sont confondues ou encore que 1 et ABC sont des *alias*.

De même on a les alias A et BC , B et AC , C et AB .

Le tableau suivant donne les estimations réellement faites :

estimation visée	estimation faite
$e(1)$	$e(1) + e(ABC)$
$e(A)$	$e(A) + e(BC)$
$e(B)$	$e(B) + e(AC)$
$e(C)$	$e(C) + e(AB)$

Dans un plan fractionnaire les actions confondues (ou aliassées) sont indissociables. La réduction du nombre d'expériences a un prix : moindre précision des estimateurs, estimation non des actions mais des effets des alias d'actions.

4. Comparaison d'un mauvais plan et d'un bon

Supposons que l'on conduit l'expérimentation d'une manière traditionnelle en faisant varier successivement chacun des facteurs. On obtient le plan suivant :

n° d'expérience	A	B	C
1	-1	-1	-1
2	1	-1	-1
3	1	1	-1
4	1	1	1

Certes, on peut estimer $e(A)$ par $\frac{1}{2}(y_2 - y_1)$

$$e(B) \text{ par } \frac{1}{2}(y_3 - y_2)$$

$$e(C) \text{ par } \frac{1}{2}(y_4 - y_3)$$

On remarquera que ces estimations ont pour écart type $\frac{\sigma}{\sqrt{2}}$ au lieu de $\frac{\sigma}{2}$ dans le cas du plan fractionnaire précédent ; elles sont donc moins précises. De plus en cherchant les niveaux des interactions, on verra qu'il n'est pas possible de trouver des alias. Résultat quand le modèle postulant que les interactions sont nulles n'est pas vérifié, on ne sait pas ce qu'on estime. Un exemple de plus pour se convaincre de la nécessité de faire des plans d'expérience.

VII. Pour aller plus loin

L'objectif de l'exposé ci-dessus était de faire connaître à partir d'un exemple simple – trois facteurs chacun à deux niveaux – les concepts essentiels des plans d'expérience : définition des actions, modèle, randomisation, estimation optimale, alias. A partir de là, on pourra explorer le vaste domaine des plans d'expérience : plusieurs facteurs à plusieurs niveaux, facteurs variant continûment, modèles linéaire ou quadratique, que faire quand certaines expériences sont impossibles (risque d'explosion par exemple) etc.