Plans d’expérience pour mélange de mélanges

Hanna Hanen, Walter Tinsson

Plans d'expérience pour mélange de mélanges

Hanen Hanna & Walter Tinsson

Université de Pau et des Pays de l'Adour, Laboratoire Mathématiques Appliquées de PAU (LMAP), UMR 5142. Avenue de l'Université, 64000 Pau, France

hanen.hanna@etud.univ-pau.fr, walter.tinsson@univ-pau.fr

Résumé

On considère ici des plans pour mélanges de mélanges, dits MoM (Mixture of Mixtures), où chaque composant principal est lui même un mélange de composants secondaires. Les contraintes de ce type de système conduisent très souvent à la sur-paramétrisation de la réponse lorsque un modèle polynomial classique est adopté, c'est pourquoi un type spécial de modélisation est requise. Les modèles proposés dans la bibliographie sont obtenus par multiplication de polynômes associés à chaque contrainte MoM. Ils sont ensuite ajustés sur des plans dit réseaux de Scheffé multiples mais leur taille est souvent problématique à mettre en oeuvre. Nous proposons alors l'adaptation des outils de la géométrie algébrique tels que les bases de Gröbner afin d'obtenir l'ensemble maximal de monômes identifiables à partir d'un plan d'expérience proposé.

Mots clés : Plan d'expérience pour mélange, mélange de mélanges, composants principaux et secondaires, modèles identifiables, bases de Gröbner.

Summary

We consider mixture-of-mixtures (MoM) experiments. In such case every major component is a mixture of some minor components. Classical polynomial models are surparametrized for MoM. One usual way to obtain regular models is to mulitply all the polynomial mixture models associated to each major component. Such models can be fitted using multiple-lattice designs but these designs are constituted by a big number of experiments. Our goal is to use algebraic geometry, and more precisely Gröbner basis, in order to obtain for every fixed design the maximal set of identifiable monomials under the MoM constraints.

Keywords : Experimental designs for mixtures, mixture of mixtures, major and minor components, identifiable models, Gröbner bases.

Système pour mélange de mélanges

En général un plan d'expérience pour mélanges est un type particulier de plan dont les facteurs expliquant certaines propriétés du mélange sont les ingrédients qui le composent. On suppose que la réponse à modéliser dépend uniquement des proportions des composants.

Un mélange classique de q composants, avec x_i la proportion du composant i, est défini par le point $(x_1, x_2, \cdots, x_q)$ de $\mathbb{R}^q$ satisfaisant les contraintes suivantes :

$$x_1 + x_2 + ... + x_q = 1 \quad et \quad \forall\ i = 1, \cdots, q, \quad x_i \geq 0$$

Les contraintes ci-dessus délimitent le domaine expérimental, qui est un polyèdre de dimension $q-1$ de $\mathbb{R}^q$. Il peut être exprimé comme l'enveloppe convexe de la base canonique de $\mathbb{R}^q$, appelé simplexe et dénoté par S_q.

Par exemple, l'ensemble de tous les mélanges à 3 composants est le triangle équilatéral dans $\mathbb{R}^3$ défini par :

$$S_3 = \{(x, y, z) \in \mathbb{R}^3 : \quad x + y + z = 1, \quad x, y, z \geq 0\}$$

On peut trouver dans le livre de Cornell (2002) des plans dits *réseaux de Scheffé* utilisables pour estimer des surfaces de réponse polynomiales. Pour m un entier positif, le $\{q, m\}-$réseau de Scheffé à q composants est le plan constitué par l'ensemble de points $(x_1, \cdots, x_q) \in S_q$ où $x_i \in \{0, \frac{1}{m}, \frac{2}{m}, \cdots, \frac{m-1}{m}, 1\}$.

Les plans pour mélanges classiques ne permettent que l'exploration d'un seul mélange à la fois. Il y a des cas où l'expérience pour mélanges est conduite afin de modéliser les propriétés de plusieurs types de mélange en même temps, donc on suppose que la formulation est un *mélange de mélanges* que l'on appellera système MoM (mixture of mixtures).

Considérons un système MoM dont les composants *principaux* (CP) sont des mélanges de composants *secondaires* (CS), chacun d'eux étant associé à un seul CP. Soit p le nombre de CP dont les proportions sont données par le vecteur $(w_1, w_2, \cdots, w_p)$ et q_i le nombre de CS associés au i-ème CP dont les proportions sont $(y_{i1}, y_{i2}, \cdots, y_{iq_i})$. Alors, le système de contraintes que lie les proportions est représenté par :

$$\forall\ \ i = 1, \cdots, p\ ;\ \forall\ j = 1, \cdots, q_i$$

$$w_i = y_{i1} + y_{i2} + \cdots + y_{iq_i} \tag{1}$$

$$w_1 + \cdots + w_p = 1 \tag{2}$$

$$w_i > 0, \quad y_{ij} \geq 0 \tag{3}$$

Piepel (1999) introduit la classification des systèmes MoM en deux types, si les proportions de CP sont fixées on dit que le plan est de type A et si elles sont variables on dit type B.

Pour les plans de type A on divise chaque équation de (1) par w_i, en définissant $x_{ij} = \dfrac{y_{ij}}{w_i}$, donc le système (1)-(3) est remplacé par les contraintes $x_{i1} + x_{i2} + \cdots + x_{iq_i} = 1$ et $x_{ij} \geq 0$.

Modèles polynomiaux

Les plans d'expérience MoM trouvés dans la littérature sont les plans proposés par Lambrakis (1968), réutilisés par Piepel (1999) et Cornell & Ramsey (1998). Ce type de plan est construit sur des réseaux multiples de Scheffé avec un modèle de réponse formé par la multiplication de polynômes *canoniques* associés à chaque contrainte (1)-(2). On appelle canonique un polynôme identifiable sur le réseau de Scheffé. Pour les degrés 1,2 et 3 ces polynômes sont obtenus par réparametrisation en remplaçant la contrainte $x_1 + \cdots + x_q = 1$ dans le polynôme usuel de même degré.

Les réseaux de Scheffé, simples et multiples, sont des plans saturés par rapport au polynôme canonique associé au plan. On dit que le plan est saturé lorsque la taille du plan est égale au nombre de paramètres à estimer.

Exemple. Pour le système MoM de type A avec $p = 2$, $q_1 = 2$, $q_2 = 3$ et les polynômes canoniques de degrés deux et un respectivement, le modèle multiplicatif adopte la forme suivante :

$$\begin{aligned} P(\mathbf{x}) = \quad & a_1x_{11}x_{21} + a_2x_{11}x_{22} + a_3x_{11}x_{23} + \\ & a_4x_{12}x_{21} + a_5x_{12}x_{22} + a_6x_{12}x_{23} + \\ & a_7x_{11}x_{12}x_{21} + a_8x_{11}x_{12}x_{22} + a_9x_{11}x_{12}x_{23} \end{aligned}$$

Bien que le modèle multiplicatif soit le plus recommandé pour étudier l'ensemble des interactions, la taille du plan requis pose des problèmes pour des nombreux CP et CS.

Très peu de modèles alternatifs existent dans la bibliographie. Nigam (1973) propose un modèle réduit à partir du polynôme quadratique classique sur deux groupes de variables secondaires traités comme des facteurs et non pas comme de composants principaux dans un mélange.

D'autre part, si l'on essaie de modéliser la réponse pour un système MoM de type A à l'aide d'un polynôme de premier ordre tel que $P(x) = a_o + \sum_{i=1}^{p} \sum_{j=1}^{q_i} a_{ij}x_{ij}$, le modèle ne sera pas identifiable à cause des contraintes $x_{i1} + x_{i2} + \cdots + x_{iq_i} = 1$, $\forall\ i = 1, \cdots p$.

Vu les difficultés entrainées par les contraintes MoM lorsqu'un modèle préconçu est adopté, il est possible d'utiliser les travaux de Pistone et Wynn (1996). Les auteurs utilisent la théorie des bases de Gröbner pour obtenir des modèles polynomiaux identifiables à partir d'un plan d'expérience donné qui est alors la solution d'un système d'équations.

L'idée de la méthode est de sélectionner un plan d'expérience de taille adéquate, de construire un système d'équations polynomiales de la forme $P_1 = 0, P_2 = 0, \cdots, P_k = 0$ (dit variété affine, dont les solutions sont les points du plan) puis d'obtenir la base de Gröbner pour l'ensemble de polynômes $\{P_1, P_2, \cdots, P_k\}$ sur un ordre choisi. À partir de la base de Gröbner on déduit les confusions d'effets et les termes dominants selon l'ordre choisi, on obtient ensuite l'ensemble de monômes identifiables pour le plan donné. On remarque que le nombre de mônomes obtenus est toujours égal à la taille du plan.

Exemple. Pour le système MoM de l'exemple précèdent, considérons le réseau multiple de Scheffé $\{q_1 = 2, m_1 = 2\}_{\mathrm{x}} \{q_2 = 3, m_2 = 1\}$ où la matrice du plan est :

$$D = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 1 \end{bmatrix}$$

Les contraintes MoM associées sont : $x_{11} + x_{12} = 1$ et $x_{21} + x_{22} + x_{23} = 1$. Pour chaque variable x_{ij} on construit une équation de la variété affine en utilisant les proportions de cette variable selon le plan donné, ainsi, pour x_{11} on a l'équation $(x_{11} - 1)(x_{11} - \frac{1}{2})x_{11} = 0$.

Le logiciel Maple peut être utilisé pour obtenir la base de Göbner et l'ensemble de monômes identifiables.

Bibliographie

[1] Cornell J. (2002) *Experiments with mixtures :designs, models and the analysis of mixture data*. 3er ed., New York : Wiley.

[2]Cornell J. and Ramsey I.(1998) *A generalized mixture model for categorized-components probles with an application to a photoresist-coating experiment*. Technometrics, 40,48-61.

[3] Lambrakis (1968) *Experiments with mixtures : A generalization of the simplex-lattice design*. Journal of the Royal Statistical Society, B,30 , 123-136.

[4] Nigam A. (1973) *Multifactor mixture experiments*. Journal of the Royal Statistical Society, B, 35, 51-66.

[5] Piepel G. (1999) *Modeling methods for mixture-of-mixtures experiments Applied to a Tablet Formulation Problem*. Pharmaceutical Development and Technology, 4, 593-606.

[6] Pistone G., Wynn H. (1996) *Generalised confounding with Gröbner bases*. Biometrika, 83, 653-666.